

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 15:00:31 ; Search time 94 Seconds  
(without alignments)  
2524.674 Million cell updates/sec

Title: US-09-995-749A-2

Perfect score: 9424

Sequence: 1 MEIKKFKLYKSKQWVTRA.....SDAEYPTSTDVGMIDQNKI 1781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :				A.Geneseq.l01002.*			
1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*						
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*						
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*						
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*						
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*						
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*						
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*						
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*						
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*						
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*						
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*						
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*						
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*						
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*						
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*						
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*						
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*						
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*						
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*						
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*						
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*						
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*						
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9424	100.0	1781	AAU74519	Lactobacillus reut
2	2724	28.9	2057	AAU10667	L. mesenteroides a
3	2574.5	27.3	1527	AAU80055	Leuconostoc mesent
4	2458	26.1	1430	AAU98044	S. mutans glucosyl
5	2454	26.0	1430	AAU98045	S. mutans glucosyl
6	2453	26.0	1430	AAU98041	S. mutans glucosyl
7	2451	26.0	1430	AAU98043	S. mutans glucosyl
8	2449	26.0	1430	AAU98042	S. mutans glucosyl
9	2446	26.0	1430	AAU98029	S. mutans glucosyl
10	2347.5	24.9	1375	23	AAU98028

11	2347.5	24.9	1375	23	AAU79288	Streptococcus muta
12	2330.5	24.7	1476	23	AAU79284	Streptococcus muta
13	2282	24.2	1475	23	AAU98033	S. mutans glucosyl
14	2280	24.2	1475	23	AAU98040	S. mutans glucosyl
15	2277	24.2	1475	23	AAU98030	S. mutans glucosyl
16	2276	24.2	1475	23	AAU98027	S. mutans glucosyl
17	2274	24.1	1475	23	AAU98037	S. mutans glucosyl
18	2274	24.1	1475	23	AAU98038	S. mutans GTFB mut
19	2274	24.1	1475	14	AAU32925	Glucosyltransferase
20	2271	24.1	1475	23	AAU98031	S. mutans glucosyl
21	2270	24.1	1475	23	AAU98035	S. mutans glucosyl
22	2269	24.1	1475	23	AAU98032	S. mutans glucosyl
23	2268	24.1	1475	23	AAU98036	S. mutans glucosyl
24	2264	24.0	1475	23	AAU98034	S. mutans glucosyl
25	2258	24.0	1475	23	AAU98039	S. mutans glucosyl
26	2190.5	23.2	1577	17	AAU91047	Alpha-D-glucosyltr
27	2183	23.2	1017	23	AAU79285	Streptococcus muta
28	344.5	3.7	5795	22	AAU37017	Staphylococcus aur
29	327.5	3.5	2411	21	AAU23860	Haemophilus influe
30	326.5	3.5	349	22	AAU26877	Glucosyltransferas
31	325.5	3.5	2314	22	AAU69136	M. catarrhalis les
32	321.5	3.4	2353	17	AAU93393	Haemophilus adhesi
33	318.5	3.4	6281	22	AAU37403	Staphylococcus aur
34	313	3.3	2902	22	AAU46351	H. pylori HP165 p
35	313	3.3	5024	22	AAU82935	S. epidermidis ope
36	311.5	3.3	1612	23	AAU47333	Listeria monocytog
37	309	3.3	2893	19	AAU98828	H. pylori GHPO 148
38	309	3.3	2893	19	AAU71556	Helicobacter polyp
39	304.5	3.2	3158	22	AAU37018	Staphylococcus aur
40	303	3.2	10182	23	ABP38314	Staphylococcus epi
41	297.5	3.2	2434	22	AAU34339	Staphylococcus aur
42	297	3.2	2500	21	AAU18272	Plasmodium falcipla
43	294.5	3.1	2086	22	AAU34143	Staphylococcus aur
44	292	3.1	3696	23	ABP40235	Staphylococcus epi
45	290.5	3.1	2659	23	AAU75489	S. aureus antigeni

ALIGNMENTS

RESULT 1  
AAU74519  
ID AAU74519 standard; Protein; 1781 AA.

AC AAU74519;

XX AAU74519;

DT 09-APR-2002 (first entry)

DE Lactobacillus reuteri glucosyltransferase A (gtfa) polypeptide.

DE Glucosyltransferase A; gtfa; glucan: anhydroglucose; sucrose;

KW oligosaccharide; polysaccharide; sucrose; fructan; symbiotic; probiotic.

XX Lactobacillus reuteri.

OS Lactobacillus reuteri.

XX WO200190372-A1.

PN 29-NOV-2001.

XX 23-MAY-2001; 2001WO-NL00393;

XX 25-MAY-2000; 2000EP-0201871.

XX (NEDE ) NEDERLANDSE ORG TOEGEPAST.

XX Van Geel Schutten GH, Dijkhuizen L, Rahaoui H, Leer RJ;

XX WPI; 2002-147583/19.

XX N-PSDB; AAS20540.

XX Protein with glucosyltransferase activity derived from Lactobacillus

PT Reuteri -

XX

Claim 3; Page 48-55; 59pp; English.

The invention relates to a Lactobacillus reuteri glucosyltransferase A (gtfa) polypeptide. This polypeptide produces a glucan with a unique structure having 4-linked, 6-linked and 4-6-linked anhydroglucose units or, in the presence of suitable acceptors, oligosaccharides. The protein can be used to produce an oligosaccharide or polysaccharide of interest, for example, to produce a glucan with sucrose as a substrate and optionally a fructan as a probiotic or symbiotic. This sequence represents *L. reuteri* glucosyltransferase A.

Sequence 1781 AA;

Query Match 100.0%; Score 9424; DB 23; Length 1781;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEIKKFKLYKSGKWVTAATVAVTALLYGGVAHADQQVOASTQDTSTVNNDTD 60  
| | | | |  
DB 1 MEIKKFKLYKSGKWVTAATVAVTALLYGGVAHADQQVOASTQDTSTVNNDTD 60  
| | | | |  
QY 61 KTVALDNTDQSAQTDDKKQVVSNINOSKTDDTSTADKNSTSPSVLPSPNTEKQAKNY 120  
| | | | |  
DB 61 KTVALDNTDQSAQTDDKKQVVSNINOSKTDDTSTADKNSTSPSVLPSPNTEKQAKNY 120  
| | | | |  
QY 121 NEODKNGYNIDTAFYSNNQLHVSQWNAATNASQGTNSRQIIVRDITTNELGRTDVNNV 180  
| | | | |  
DB 121 NEODKNGYNIDTAFYSNNQLHVSQWNAATNASQGTNSRQIIVRDITTNELGRTDVNNV 180  
| | | | |  
QY 181 ARPDKNVHNVDNADNSGFVNVDNIDFSKMKYDRSDIEISRYSGNGSKSYDMWSQPITED 240  
| | | | |  
DB 181 ARPDKNVHNVDNADNSGFVNVDNIDFSKMKYDRSDIEISRYSGNGSKSYDMWSQPITED 240  
| | | | |  
QY 241 KNNAYLDTPEYKNGELHATGNATNSAINNNHHFVILFDOTNGKEVAROEVEGQS RPD 300  
| | | | |  
DB 241 KNNAYLDTPEYKNGELHATGNATNSAINNNHHFVILFDOTNGKEVAROEVEGQS RPD 300  
| | | | |  
QY 301 VAKVYPQVYGAANS GFNFTEFNISDLDYTHQYQVLRSYNSDNGBGDNVTWFNPQS IAPA 360  
| | | | |  
DB 301 VAKVYPQVYGAANS GFNFTEFNISDLDYTHQYQVLRSYNSDNGBGDNVTWFNPQS IAPA 360  
| | | | |  
QY 361 NQSNQGLDSDFDI SKNGEVTYTGWNATDSELQNHHYVILFDQTAGQVASAKADLI SRP 420  
| | | | |  
DB 361 NQSNQGLDSDFDI SKNGEVTYTGWNATDSELQNHHYVILFDQTAGQVASAKADLI SRP 420  
| | | | |  
QY 421 DVAKAYPTVKTTATNSGFKYTFKVNNLQPGHQYVYVSRFSADENGNGNDRKHTDYWFSPVI 480  
| | | | |  
DB 421 DVAKAYPTVKTTATNSGFKYTFKVNNLQPGHQYVYVSRFSADENGNGNDRKHTDYWFSPVI 480  
| | | | |  
QY 481 LNQTASNIDTIWTNSGLHIAGWASDNSINETTPYAIIILNNGKEVTRQKMSLTARPDVA 540  
| | | | |  
DB 481 LNQTASNIDTIWTNSGLHIAGWASDNSINETTPYAIIILNNGKEVTRQKMSLTARPDVA 540  
| | | | |  
QY 541 AVTPSYLSINAVSGFDFTTKLTNDQYQALNGOLVLLRFPSKAADGNPSGDNTVTDQFSKNY 600  
| | | | |  
DB 541 AVTPSYLSINAVSGFDFTTKLTNDQYQALNGOLVLLRFPSKAADGNPSGDNTVTDQFSKNY 600  
| | | | |  
QY 601 ATTGGNFDDVYKNGNQVEFGSHWATNQSDNKDSOWIIVLVNKGKEVKQLVNDTKEGAAGF 660  
| | | | |  
DB 601 ATTGGNFDDVYKNGNQVEFGSHWATNQSDNKDSOWIIVLVNKGKEVKQLVNDTKEGAAGF 660  
| | | | |  
QY 661 NRNDVYKVNPAIENSMSMGFQGIITLPVTVKNNVQLVHRFSNDVKYTGEGNYVDFWSELM 720  
| | | | |  
DB 661 NRNDVYKVNPAIENSMSMGFQGIITLPVTVKNNVQLVHRFSNDVKYTGEGNYVDFWSELM 720  
| | | | |  
QY 721 PVKDSFOKNGPLKQFGLQTINGQOYYIDPTTGGOPRKNFLLOSNNWIYFDSDTGVGTNA 780  
| | | | |  
DB 721 PVKDSFOKNGPLKQFGLQTINGQOYYIDPTTGGOPRKNFLLOSNNWIYFDSDTGVGTNA 780  
| | | | |  
QY 781 LELOFAKGTVSSNEQYRNQNAAYS YDDKS IENVNGYLTADTWYRPKOILKDGTWTDSKE 840  
| | | | |  
DB 781 LELOFAKGTVSSNEQYRNQNAAYS YDDKS IENVNGYLTADTWYRPKOILKDGTWTDSKE 840  
| | | | |

DE L. mesenteroides alternan sucrose protein.  
 KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.  
 XX  
 OS Leuconostoc mesenteroides.  
 XX  
 PN DE19905069-A1.  
 XX  
 PD 10-AUG-2000.  
 XX  
 XX 08-FEB-1999; 99DE-1005069.  
 XX  
 XX 08-FEB-1999; 99DE-1005069.  
 XX  
 XX (PLAN-) PLANTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX  
 XX Kossmann J, Welsh T, Quanz M, Knuth K;  
 XX  
 XX WPI; 2000-550294/51.  
 DR N-PSDB; AAA97904.  
 XX  
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production -  
 XX  
 XX Claim 1a: Page 30-36; 64pp; German.  
 XX  
 CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 XX  
 XX Sequence 2057 AA;  
 QY  
 Query Match 28.9%; Score 2724; DB 21; Length 2057;  
 Best Local Similarity 36.8%; Pred. No. 2.8e-143;  
 Matches 677; Conservative 210; Mismatches 453; Indels 500; Gaps 56;  
 QY 8 KLYKSGKQVTAAY--ATVAVSTALLYGGVAHADQVQOAASTQDOTSTVNNNDIKTVAL 65  
 DB 11 KLYKSGKQVTAAYTAFAVLGVSTVT---TVHADTNSVAVKQINNTGTNDSGEKKVPVP 66  
 QY 66 DTNTDQSAQTTPDKQVVSNTNQ--SKTDP---TSTADKNSTSTPVSLPSNNTXEOAKNY 120  
 DB 67 STNDSLKQGTDFGWYDSGDNRVQDKTQNTLLTAEOLEKKNNEKNLSVSDDTSKKDDENI 126  
 QY 121 NEQDK-GNYCNIDTAFNSNQLHVSNNATNASQGTNSQIIIVRDTTNNELGRDVTNN 179  
 DB 127 SKOTIKANQOYDTA-----KGLTTSN-----LSDP 152  
 QY 180 VARPDKVKNHNV--NADNSGFDVN--VNIDFSKMKDYRDSIEIVSRYSNGSKSVDMWSQ 235  
 DB 153 ITGGHYEN-HNGYFVIDASGKQVTLQNI-----GNLQ----- 186  
 QY 236 PITFDKNNYAYLDTFEVKNGELHATGWNATNSAINTNHHFVLFQOTNGKEVARQEVRRG 295  
 DB 187 --YFDDNGYQVKSFRDVNGK-----H--IYFDSVTGKASSNVDIVNG 225  
 QY 296 QSRPDKVAKYPPQVVGAAANGSGFNTEISDLDTHQVQLSRYSNDSNGEGDNVYWFNPQ 355  
 DB 226 KAQ-----GYDAQGN-----QLKKSIVADSSGQ-----TYIFD-- 253

QY 356 SIAPANOSNOGYLDSFDISKNGEVTVTGNATDLSLONNHYYVILFDQTAGKQVASAKAD 415  
 DB 254 -----GNGQ----- 257  
 QY 416 LISRPDVAKAYPTVKATNSGFKVTFKVNLPQGHQYVSVSRFSADENGNDKRHTDYW 475  
 DB 258 -----PLIGLQITDG-----NLQYFNQOQVQIK-----GGFQDVNNKRIY 292  
 QY 476 FSPVILNOTASNIDTTMTNSGLHIAGWMAASDNSINETTPYAILLNGKEVTPQKMSLIA 535  
 DB 293 FAP-----NTGNAVANTEI----- 306  
 QY 536 RPDVAAYPSLYNSAVSGFDTTIKLTINDQYQALNGQLQVLLRFESKAADGNPSGDNVTVDQ 595  
 DB 307 -----INGKLQ-----GRDANGNOVKNA 324  
 QY 596 FSKNYATTGNEDYKVVNGNOVEFSGWHATNQSNDKDSQWIIIVVNGKEVKROLVNDTKE 655  
 DB 325 FSKDVA---GNTFEDANG----- 340  
 QY 656 GAAGFENDYKVNPAJENSMSGFGIITLPVTVKNNVOLVHRFESNDYKTGEGNYVDF 715  
 DB 341 -----VMLT----- 344  
 QY 716 WSELMPVKDSFQKNGPLKQFGLQTINGOQYYIDPTTGQPRKNFLQSGNNWYIFSDDTG 775  
 DB 345 -----GLQTSIGKTYILD-EOGHLRKNYAGTFNQFYFDADTG 382  
 QY 776 VGTNALBLOFAKQVSSNEQYRNGNAAYSDDKSIENVNGYLTADTWYRKQILKDGTTW 835  
 DB 383 AGKTAIYQDQGLVLSNSNTPHNAAKSKDSFENVDSYLTADTWYRTDILKNGDTW 442  
 QY 836 TDSKETDMRPILMWWPNTLTQAYLYNMKQHGNNLPSALFFFNADADPAELNNHYSIVQ 895  
 DB 443 TASTETDMRPLLTWPDKQTAQYLNEMSSKGLGIITT---YTAATSQKTLNDAARVQ 499  
 QY 896 QNIKRISETGNTQWRLTHEDFVTPNPMKNKDSNVNFSGIQF-QGGFLKY-ENSLDLP 953  
 DB 500 TAIQQQSLKKSLEWLRDAIDSFVKTQANWNKOTEDSAFDGLQWLQGLAYQDSDHRT 559  
 QY 954 YANS-DYRLLGRMPINI---KDOT-VRGQBFLLANDIDNSNPVVAQBLNWLYYLLNFGT 1008  
 DB 560 NTDSSNNRKLGRQPINIDGSKDTTDDGKSEFLLANDIDNSNPVVAQBLNWLHYLMNFGS 619  
 QY 1009 ITANNDQANFDSVRYPADPNIDADLMNIAQDYFNAAQMD-SDAVSNKHTNILEDNHAD 1067  
 DB 620 ITGNNDNANFQIRVDVADVNDADLLKTAGDYFRALYGTDKSDANANKHLSILEDWNGKD 679  
 QY 1068 PEYFNKIGNPOLTMDDTIK---NSLNHGLSDATNRW-----GLDAIVHQSLADREN- 1115  
 DB 680 PQYVNOQNAQLTMDYTVTSQFNSLTHGANNRSMNYFLDTGYLLNGDLNKKIVDKNRP 739  
 QY 1116 -----NSTENVVIPNYSFVRAHDNNSQDOIQNAIRD--VTGKDYHTFTFEDBQK 1163  
 DB 740 NSGTLVNRANSQGTVPNYSFVRAHDYDAQDPIRKAMIDHGIIKKNQDFTTFDQLAQG 799  
 QY 1164 IDAYIQDN--STVKKYNLYNPASAYLLTNKDTIPRYYYGDLTYDGGQYMEHQTRYD 1221  
 DB 800 MEFYKQDENSGFKKYNDYLNPSAYAMLTKNDTPRVYYGDMYLEGGQYMEKGTIYNP 859  
 QY 1222 TLTNLLKSRVYVAGGQSNQTMVSQ-----GNNNLTISVRYGKGMATYDGTDET--- 1272  
 DB 860 VISALLKARIKVVSGQTMATDSSGKLDGDETDLTTSVRFKGMTSDQTQTDNSODY 919  
 QY 1273 RFOGIVVSVTPNLKLGVDNKKVVLHMGAAHKNQYRAAVLTITTDGVNLTNSDQAPVAM 1332  
 DB 920 KNOGLGVIVGNPDPLKLNNDKTIITLHMGAHKNQYRALVLSNDSGIDVYDSDOKAPTLR 979  
 QY 1333 TDENGDLYLSSHNLYVNGKEA---DTAVQGYANDVSGYLVAVVYVPGASNDQARTAPS 1389  
 DB 980 TNDNGDLIFHKINTEV--KQDGTIINYEKMSLNALISGYLGWVYVPGASDQARTV-A 1036



Db 842 TADGLAYDTDENAPVAYTANGDLIFTNE-----SIYGVQNPQVSGYLAVVW 889

Qy 1375 PVGASNDQARTAPSEKNSGNSAYRTNAAFDSNVIFEAFSEFVYPTKRESEANVRIAQ 1434

Db 890 PVGQQQDQARTASDTTNTSDVFNHSAALDSQVIEGFSNFQAFATSDSESYTNVVIQA 949

Qy 1435 NADFFASLGFSTFEMAPQYNSKDRFLDSTIDNGYAFTDRYDLGMSHPNKYGTDEDLRN 1494

Db 950 NADQFKWGVTSFQLAPAFYSRSDTSFLDSIIQNGYAFTDRYDLGVTPTKYGTADQLDR 1009

Qy 1495 AIGALHKAQLQVWADWVPDQIYNLPCKEVATVTRVDDRGVNWDAIINNLLYVNTIGGG 1554

Db 1010 AIGALHASKGQAADWVPDQIYNLPQELATVTRVNSFGDDQVDSQIDNALYVYVQGRGG 1069

Qy 1555 EYQKYGGAFLDLQKLYPBIFFKQVSTGVAIDPDSOKITEWSAKYFNFTNHLRSGSYV 1614

Db 1070 QYQFMYGGAFLDLQKLYPBIFFKQVSTGVAIDPDSOKITEWSAKYFNFTNHLRSGSYV 1129

Qy 1615 LKADG-QQYINLGLTTK--QFLPTLTGKKGQNEGFVKGNDGNYFYDLAGHNVKNTFI 1671

Db 1130 LKMGSKNYFKVNSNTEDGDLPRQUTNDLSE--TGFTHDDKGIY-YTLGSGYRAGNAFI 1186

Qy 1672 EDSVGNMYFFDQDGKMYE-----NKHFVDYDSYGEKGTFFFLKNGYSFRGLVQTDNGT- 1725

Db 1187 QDDNNYFFDKTGLHVTGLQKINN------TYFELPGLIELYKSFQNEGDGTI 1235

Qy 1726 YFPDNYGMYRNQTI---NAGAMYITLDENGLIK---ASVNSDAEYPTSTDYGVKMLDQ 1779

Db 1236 YFPDKKHGVFDQYITDONGNA--YFPDAGVMLKSLGATIDGHQY-----FDQN 1284

RESULT 4

AAU98044

ID AAU98044 standard; Protein; 1430 AA.

XX AAU98044;

XX 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFD mutant N471D/T589D.

KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 471

FT /note= "Wild-type Asn substituted by Asp"

FT Misc-difference 589

FT /note= "Wild-type Thr substituted by Asp"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-0740274.

XX 11-DEC-1998; 98US-0210361.

PR 07-JUN-1995; 95US-0478704.

PR 07-JUN-1995; 95US-0482711.

PR 07-JUN-1995; 95US-0485243.

PR 16-JAN-1998; 98US-0007999.

PR 16-JAN-1998; 98US-0008172.

PR 20-JAN-1998; 98US-0009620.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

DR

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in

PT paper manufacture, comprises mutations in specific positions -

XX Claim 36; Page -: 44pp; English.

PS The invention an isolated protein comprising a glucosyltransferase

CC (GTF) B polypeptide having changes at position from I448V, D457N,

CC D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K,

CC D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T,

CC Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having

CC changes at positions from T589D, T589E, N471D, N471D/T589D, and

CC N471D/T589E. Also included are a glucan produced by the GTF mutant,

CC an isolated polynucleotide which encodes P1 or P2, or its complementary

CC polynucleotide, a ribonucleic acid sequence encoding the GTF mutant,

CC an expression cassette comprising the polynucleotide operably linked to a

CC promoter, a vector comprising the expression cassette, host cell

CC introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilising the glucan produced by GTF, which utilises

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step.

CC The present sequence represents a GTFD mutant of the invention.

CC Note: The present sequence is not shown in the specification but

CC was created by the indexer using the GTFD sequence appearing as AAU98029

CC and the information in claim 36.

XX Sequence 1430 AA;

Query Match 26.1%; Score 2458; DB 23; Length 1430;

Best Local Similarity 46.9%; Fred. No. 1.2e-126;

Matches 510; Conservative 169; Mismatches 309; Indels 100; Gaps 25;

QY 738 LQINGQQYYIDPTTQGPKNFLLQSGNNWYIFDSQYGTNALELQFAGKTVSSNEQYR 797

Db 174 VKQIDGKYYVIG-SDGQPKKFNALTVNNKLVYFDKNTGALTDTQYQFQGLTKLNNDYT 232

QY 798 NGNAAYSYDDKSIENVNGYLTADTWYRPQILKDGTTWTDSDKTEMRPLMWWPNTLQ 857

Db 233 PHNQIYVNFNTSLETIDNVYTDASWYRPQILKNGTKTWTASSEDRLRPLMWSWPDQYQ 292

QY 858 AYYLNYMKQHG-----NLLPSALPFFNADAPALNHYSEIVQONIEKRISSETGNDWL 912

Db 293 IAYLNYMNOOGLGTGEN-----YTDSSQESLAAQTVQVKIETKISOTQOTOWLR 344

QY 913 TLMHDFVYNNPMKDSNVNFGSIO--FCGGFLKYENSNDLTPYANSYRLLGRMPIN-- 968

Db 345 DIINSEVKTQPNWNSQTSDESISAGEKDLGGGALLYSNSDKTAYANSYRLLNRTPSOT 404

QY 969 -----IKDQTYRGQEFLLANDIDNSNPVYQAEOLNWLNYLLNFGTITANNDOANFSDV 1023

Db 405 GKPKYFEDNSSGGYDFLLANDIDNSNPVYQAEOLNWLHLMNYGSIIVANDPEANFDGVRV 464

Qy 1024 DAPDNIDADLNIAQDYFNAAYGMD--SDAVSNKHNILEDMWHDADPEYFNKIGNPQLTMD 1082  
Db 465 DAVNDVADLEQIAISDLKALHYGVDKSEKNAINHLSLEAWSDNDPOYNKDKTKGAQLPID 524  
Qy 1083 DTIKNSLNHGLS-----DATN----RGLDAIVHOSLADRENNSTENVVNPYSFVRAHD 1133  
Db 525 NKLSLLYALTRPLEKADSKNKNEIRGLPEVITNSLNNSAEGKSERMANVIFIRAH 584  
Qy 1134 NNSQDQIQNATR-DVTGK-DYHITTFDEQKIDAYIQDQNSNTVKYNLYNPASAYAILL 1191  
Db 585 SEVDQVIAKIIKAQINPKTKDTGLTDELKQAFKIYVEMRQAKKKTQSNIPTAYALML 644  
Qy 1192 TNKDTIPRVYGDLYTGGQYMEHOTRYDPLTNLLSRKYVAGGOSMOTMSVCGNNN- 1250  
Db 645 SNKDSITRLYGDWYSDGQYMARKSPYDAIDTLKARKIYAAGQDMKITVEGDKSH 704  
Qy 1251 -----ILTSVRVYKGAAMTATDGTDETRPGIGVYVSNPTNLKLVNDKVLHMGAAH 1303  
Db 705 MDWDYTGVLTSVRVGTGANEATDQGSSEATKTQGMVAVITSNPNSLKLNDKVIINMGAAH 764  
Qy 1304 KNGQYRAAVLTGTVGNYTSDQCAP--VAMTDENGDIYLSHNLVYNGKEADTAVQGY 1361  
Db 765 KNGEYRPLLTKDGLTSYSDAAKSLYRKTNDKGELVFEDASD-----IQGY 812  
Qy 1362 ANPDVSGYLAIVVPGASDNDARTAPSTKNSGNSAYRTNAAPDSNVIFEAFSNFYVTP 1421  
Db 813 LNPQVSGYLAIVVPGASDNDQVRAASNKANATGQVYESSALDSQLIYEGFSNFDFV 872  
Qy 1422 TKESERANVIAQNAADFASLGFTSFEMAPQYNSKRTFLDSTIDNGYAFTRDYDLGMS 1481  
Db 873 TKDSDYTNKKIAQNVQLFKSGVTSFEMAPQYSSDPSGLDSIIQNGYAFEDRYDLAMS 932  
Qy 1482 EPNKYGTDEDLRNAIQALHAKAGLOVMAADWPDQIYNLPGRVAVTVRVDGRGNVWKDAII 1541  
Db 933 KNNYGSQDMINAVKALHKSQIOVIADWPDQIYNLPGRVAVTVRVDYGEYRKDSEI 992  
Qy 1542 NNNLYVNTVITGGG-EYOKKYGCAFLDKLQKLYPEIFTKKQVSTGVAIDPSQKITEWSAKY 1600  
Db 993 KNTLYAANTSKNGDKYQAKYGGAFSELAARYPISFIRNTQISNGKKIDPSEKITAWRAKY 1652  
Qy 1601 FNGTNILHRSGYVYLKADGG-QYVNL-GTTTKQFLPTQLGCKKQNGEFGVKGNDGN-YY 1657  
Db 1053 FNGTNILHRSGYVYLKONASDKYFELKGNQT--YLPQMT--NKEASTGEV--NDGNGMT 1106  
Qy 1658 FYDLAGNNKVNFTIEDSVGNWYFFDQCKMYENKHFDVDSYGEKGYTFYFLKNGVSPRGG 1717  
Db 1107 FYSTSGYQAKNSFQDQAKGNWYFFDNGHMYVGLQQLN---GF--VQYFSLNGVQLRES 1160  
Qy 1718 LVQI-----DNGTYIEDNYGK-----MVRNQTINAGAMITYTLDEN 1752  
Db 1161 FLENADGSKNFGHLGNRYSGYISFDNSKWRFYDASGVAVGLKTINGNTQYF--DQD 1218  
Qy 1753 GKLKASY 1760  
Db 1219 GYQVKGAW 1226  
RESULT 5  
AAU98045  
ID AAU98045 standard; Protein; 1430 AA.  
XX AC AAU98045;  
XX DT 27-AUG-2002 (first entry)  
XX DE S. mutans glucosyltransferase GTFD mutant N471D/T589E.  
XX KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
KW coating composition; glucan; starch; latex; thermoplastic molecule;  
KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
XX OS Streptococcus mutans.  
OS Synthetic.

XX Key Location/Qualifiers  
FH Misc-difference 471 /note= "Wild-type Asn substituted by Asp"  
FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"  
FT  
XX US2002031826-A1.  
XX 14-MAR-2002.  
XX 19-DEC-2000; 2000US-0740274.  
XX 11-DEC-1998; 98US-0210361.  
PR 07-JUN-1995; 95US-0478704.  
PR 07-JUN-1995; 95US-0482711.  
PR 07-JUN-1995; 95US-0485243.  
PR 16-JAN-1998; 98US-0007999.  
PR 16-JAN-1998; 98US-0008172.  
PR 20-JAN-1998; 98US-0009620.  
XX (NICH/) NICHOLS S E.  
XX Nichols SE;  
XX WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions -  
Claim 36; Page -: 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, D567T/D571K/K1014T, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFD mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFD sequence appearing as AAU98029 and the information in claim 36.

SQ	Sequence.	1430 AA;	
Query Match	26.0%; Score 2454; DB 23; Length 1430;		
Best Local Similarity	46.8%; Pred. No. 2.1e-128;		
Matches 509; Conservative 170; Mismatches 309; Indels 100; Gaps 25;			
QY	738	LOTINGQYYIDPTGQPKRKLQSGNNWYFDSDTGVGTNALEQFAKGVSSNEQYR	797
DB	174	VKQIDGKYIYG-SGQPKKFAITVNNKVYFDKNTGALTDTSOYQFKQGLTKLNNDYT	232
QY	798	NNAAYSVDXSIENYNYLADWYRKQILKDGTTWDSKETDMPILMWVWPNTLTQ	857
DB	233	PHNQIVNFENTSLTIDNYVTADSWYRKDKILKNGKTWTSSESRLPLMLSWPDKQTQ	292
QY	858	AYLVNMQHG-----NLLPSALPFENADAPAEALNHVSEIVQONKRIKRISETGNTDCLR	912
DB	293	IAYLVNMQOGLGTGEN-----YTADSSQESLNLAQTQVQVKIEKISQTQQTQWLR	344
QY	913	TLMHDFVTNNPMNKDSENVPFSGIQ--FQGGFLKYENDLTPYANSYRLLGRMPIN--	968
DB	345	DIINSEVKTQPNWNSQTESDTSAGEKDHQGLGALLYSNSDKTAVANSYRLLNRTPTSQT	404
QY	969	-----IKDQTYRGQFELANDIDNSNPVYQAEQLNWLVLNFGTITANNQANFDSVRY	1023
DB	405	GRPKYFEDNSSGGYDFELANDIDNSNPVYQAEQLNWLVLNFGTITANNQANFDSVRY	464
QY	1024	DAPNDIDALMNIADQYFNAAQMD-SDAVSNKHINILEDMNHADPEYFNKIGNFQLTMD	1082
DB	465	DAVDNVDALLQIASDYLAHKHYGVKSEKNAINHLSILEAWSNDPQYNKDKTGAQLPID	524
QY	1083	DTIKSLNHLG-----DATN-----RWGLDAIVHQSLADRENNSTENVVINYSEVRAHD	1133
DB	525	NKLRSLYALTRPLEKQASNKNEIRSGLEPVITNSLNNRSAEGKNSERMANYIFIRAH	584
QY	1134	NNSQIQONAIR-DVTGK-DYHTEFFEDQGDIDAYIQDNSTVKRYNLYNIPASYAILL	1191
DB	585	SEVQEVIAKIIKAQINPKTDGLTFTDELKQAFKTYNEDMRQAKKKTQSNIPTAYALML	644
QY	1192	TNKDIPRVYGDLYTDGQYMEHTRYDYTLNLLKSRVYAGGQSMQMSVGGNNN-	1250
DB	645	SNKDSITRIYGDMSDDQGMATSKYDAIDTLKARIKYAAGGQDKMITYVEGDKSH	704
QY	1251	-----ILTSVRYGKGAMTATDGTDETRTGIGVYNSVNPNLKLVNDKVKVLMHGAH	1303
DB	705	MDWDYTGVLTSVRYGTGANEATDQSEATKIQGMVITSNNPSLKLNDKXIVVNNGAH	764
QY	1304	KNOYRAAVLTITDGVINYTSQDQAP--VAMTDENGDIYLSHNLVYNGKEADTAVQGY	1361
DB	765	KNOYRPLLLTTKDLTSYTSDAAKSLYRKNTDKGELVFDASD-----IQGY	812
QY	1362	ANPDVSGYLAVVVPVPGASDNQDARTAPSTEKNSGNSAYRTNAAFDNSVIFAFSFFVYTP	1421
DB	813	LNQVSGYLAVVVPVPGASDNQDVRVAASNKANATQGVYESSALDSQLYBGFNSFQDFV	872
QY	1422	TKESERANVRIAQADFFASLGFTSFEMAPQYNSKDKRTFLDSTIDNGYAFTRDYDLGMS	1481
DB	873	TKDSYTNKKIAQNVQLFKSWGVTSEFAPQVYSSDSGLDSIIQNGYAFEDRYDLAMS	932
QY	1482	EPNKYGTDEDLNAIQALHAKGLQYMAWVDPQIYNLPKGEVATVTRYDDRGNVKWDALII	1541
DB	933	KNNKYGSDQDMINAVKALKSGIQIADVPDQIYNLPKGEVATVTRYVNDGEYRKDSEI	992
QY	1542	NNNLVYVNTIGGG-EYQKRYGGAFDLKQLYPEIFTRKKQVSTGVAIDPSOKITEWSAKY	1600
DB	993	KNTLYAANTKSKGDKYQAKYGGAFSELAAKYPSIFNRTQISNGKKIDPSEKITAFAWKAKY	1052
QY	1601	FNGTILHRSGSYVLKADGG-QYYNL-GTTTKQFLPIQLTGEKQNGSGFVKNGDGN-YY	1657
DB	1053	FNGTILHRSGSYVLKADGG-QYYNL-GTTTKQFLPIQLTGEKQNGSGFVKNGDGN-YY	1657
QY	1658	FYDLAGNMKNFTIEDSVGNWYFFDQDGKMWENKHFVDVDSYGEKGTFFFLKNGYSFRGG	1717
DB	1107	FYSTSGYQAKNSFVDQAKGNWYFFDNGHMYGLQQLN-----GE--VQYFLSNGVQLRES	1160
QY	1718	LVQT-----DNGTYFFDNQYK-----MYRNOTINAGAMIYTLDEN	1752
DB	1161	FLENADGSKNYFGLHGRNGYYSFSDNDSKRYFDSAGVMAYGLKTINGTQYF--DQD	1218
QY	1753	GKLIKASY 1760	
DB	1219	GYQVKGAW 1226	
RESULT 6			
AAU98041			
ID	AAU98041	standard; Protein; 1430 AA.	
XX	AAU98041;		
XX	27-AUG-2002	(first entry)	
XX	S. mutans	glucosyltransferase GTFD mutant T589D.	
DE	XX	Glucosyltransferase; GTFD; transgenic plant; paper sizing;	
KW	XX	coating composition; glucan; starch; latex; thermoplastic molecule;	
KW	XX	anyoplast; vacuole; paper manufacture; mutant; mutein.	
OS	XX	Streptococcus mutans.	
OS	XX	Synthetic.	
FT	XX	Key	
FT	XX	Misc-difference 589	
FT	XX	Location/Qualifiers	
FT	XX	/note= "Wild-type Thr substituted by Asp"	
XX	US2002031826-A1.		
PN	XX	14-MAR-2002.	
PD	XX	19-DEC-2000; 2000US-0740274.	
PF	XX	11-DEC-1998; 98US-0210361.	
PR	XX	07-JUN-1995; 95US-0478704.	
PR	XX	07-JUN-1995; 95US-0482711.	
PR	XX	16-JAN-1998; 98US-0007999.	
PR	XX	16-JAN-1998; 98US-0008172.	
PR	XX	20-JAN-1998; 98US-0009620.	
XX	(NICH/) NICHOLS S E.		
XX	Nichols SE;		
XX	WPI; 2002-414332/44.		
DR	XX	Glucosyltransferase B or D protein useful for producing a glucan useful	
PT	XX	as substitutes for and additions to modified starch and latexes in	
PT	XX	paper manufacture, comprises mutations in specific positions	
XX	XX	Claim 36; Page -; 44pp; English.	
PS	XX	The invention an isolated protein comprising a glucosyltransferase	
CC	XX	(GTF) B polypeptide having changes at position from I448V, D457N,	
CC	XX	D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K,	
CC	XX	D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T,	
CC	XX	K169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having	
CC	XX	changes at positions from T589D, T589E, N471D, N471D/T589D, and	
CC	XX	N471D/T589E. Also included are a glucan produced by the GTF mutant,	
CC	XX	an isolated polynucleotide which encodes p1 or p2, or its complementary	
CC	XX	polynucleotide, a ribonucleic acid sequence encoding the GTF mutant,	
CC	XX	an expression cassette comprising the polynucleotide operably linked to a	
CC	XX	promoter, a vector comprising the expression cassette, host cell	
CC	XX	introduced with the vector, a transgenic plant comprising the	
CC	XX	vector, a seed or tuber from the transgenic plant, a paper sizing and/or	
CC	XX	coating composition comprising a glucan produced in a plant transformed	
CC	XX	with a gene encoding the mutant GTF, wild type or, starch, a latex,	
CC	XX	thermoplastic molecule or their combinations or glucan and starch where	



CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. the method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilising the glucan produced by GTF, which utilises  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step.  
CC The present sequence represents a GTF mutant of the invention.  
CC Note: The present sequence is not shown in the specification but  
CC was created by the indexer using the GTFD sequence appearing as AAU98029  
CC and the information in claim 36.  
XX  
SQ Sequence 1430 AA;

Query Match 26.08; Score 2453; DB 23; Length 1430;  
Best Local Similarity 46.88; Pred. No. 2.4e-128;  
Matches 509; Conservative 170; Mismatches 309; Indels 100; Gaps 25;  
  
QY 738 IOTINGQOYYIDPTGQPRKNFLQSGNNWYFDSGTGVTNALELOFAKQTVSSNEQYR 797  
DB 174 VKQIDGKYIIG-SDGQPKNFALTNNKVLFDKNTGALTDTSQYQKQGLTKLNNDYT 232  
  
QY 798 NGNAAYSDDKSIENGYLADTWYRPKQILKGGTWTDSKETDMPILMWWPNTLQ 857  
DB 233 PHNQIVNFENTSLETIDNVYADSWYRPKQILKNGKTWTASSESDPLRLMSWPFQKQ 292  
  
QY 858 AYYLNYMKQHG-----NLPSALFEFFNADAPAEINHYSEIVQNIIEKRISGETNTDWL 912  
DB 293 IAYLNYMNOQGLGTGEN-----YADSSOESLNLAAQVQVKIETKISQTOQTQWL 344.  
  
QY 913 TLMHDFVTNNPWKDSNNVFSGIQ--FQGFLLKYENSDTTPYANSDYRLLRMPIN-- 968  
DB 345 DIINSFVKTPNWSQTESDTSAGEKDHQGLLYSNSDKTAYANSDYRLLRNTPTSQT 404  
  
QY 969 -----IKDOTYRGQEFFLANDIDNSPVVQAEQLNWLYLLNFQGTITANDQANFDSVRV 1023  
DB 405 GKPKYFEDNSSGGYDFLLANDIDNSPVVQAEQLNWLYLLMNYGSIIVANDEANFDSVRV 464  
  
QY 1024 DAPDNIDALNIAQDYFNAYGMD-SDAVSNKHINILEDNHNADPEYFNKIGNPQTLTMD 1082  
DB 465 DAVDNVNADLQIASDYLRKHYGVKSEKNAINHLSLEAWSDDNDPOYNKDTKGAQLPID 524  
  
QY 1083 DTIKNSLNHGLS-----DATN-----RWGLDAIVHQSLADRENNSTENVIPNYSFVRAHD 1133  
DB 525 NKRLSLIYALTRPLEKDASNKNEIRSGLEPVITNSLNNRSAEGKNSERWANYIFIRAH 584  
  
QY 1134 NNSQDQIONATR-DVTGK-DYHTTFDEQKIDAYTQDQNTVVKYNLYNPASAYAILL 1191  
DB 585 SEVQDVIKIIKAQINPKTDLTITDELKQAFKIYNEDMFQAKKYYTQSNIPYAYALML 644  
  
QY 1192 TNKDTIPRVYGDLYTDGQYMEHQTRYIDTLTLLKSRVYVAGGSMQTMVSGGNN- 1250  
DB 645 SNKDSITRLXYGDMYSDGQYMATKSPYDAIDTLLKARIKAYAGGQDMKITVEGDKSH 704  
  
QY 1251 -----TLTSVRVYKGMATATDGTDETRQG:GVVYVSNPNLKLGVNDKWLHMGAAH 1303  
DB 705 MDWDYTGVLTSVRVYGTGANEATDQGEATKTQGMNAVITSNNPSLKNQNDKVIWNMGAAH 764  
  
QY 1304 KNOQYRAAVLTITDGVINVTSDQAP--VAMTDENGDIYLSHNLVYNGKEADTAVQGY 1361  
DB 765 KNOEVRPLLLTKGLTSYTSDAAKSLYRKNTDKGLVDFDASD-----IQGY 812

QY 1362 ANPDVSGYLAVVVPVCGASDNQDARTAPSTEKNSGNSAYETNAAFDSNVIFEAFSFEVYTP 1421  
DB 813 LNPQVSGYLAVVVPVCGASDNQDVRVAASKNATAGVYESSALDSQLIYEGFSNFQDFV 872  
  
QY 1422 TKESERANVRIAQADOFFASIGFTSFEMAPQYNSSKDRITFLDSTIDNGYAFTRDYDLGMS 1481  
DB 873 TKDSYTNKKIAQNVQLFKSWGVTSEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMS 932  
  
QY 1482 EPNKYGTDEDLRNAIQALHKAGLOVADWVPDQIYNLPKGEVATVTRVDDRGVWVKDAII 1541  
DB 933 KNNKYGQSDQMINAVKALHKSQIQVADWVPDQIYNLPKGEVATVTRVNDYGEYRKDSEI 992  
  
QY 1542 NNNLYVNVYIGGG-EYQKKYGGAFDLKQRLYPEIFTKKQVSTGVAIDPSQKITEMSARY 1600  
DB 993 KNTLYAANTKSNCKDQYQAGYGAFLSELAAKYPSIFNRQISNGKKIDSEKITAMKAY 1052  
  
QY 1601 FNGFTNIIHRGSGVYLKADGG-QYVNL-GYTKQFIPQITGKKGQNEFVKGNDGN-YY 1657  
DB 1053 FNGFTNIIHRGSGVYLKADGG-QYVNL-GYTKQFIPQITGKKGQNEFVKGNDGN-YY 1657  
  
QY 1658 FYDLAGNMVKNTEFDSVGNWYFFDQDGRKVENKHFVDVDSYGEKGYFFLKNGVSFRCG 1717  
DB 1107 FYSTSGVQAKNSFVQDAKGNWYFDDNNGHMYVGLQQLN---GE--VQYFSLNGVOLRES 1160  
  
QY 1718 LVQV-----DNGTYIFDNYGK-----WVRNOTINAGAMIYTLDEN 1752  
DB 1161 FLENADGSKNYFGLHGNRYSNYISFDNDSKWRYPDASGVMAVGLKTINGNTQYF--DOD 1218  
  
QY 1753 GKLIKASY 1760  
DB 1219 GYQVKGAW 1226  
  
RESULT 7  
AAU98043  
ID AAU98043 standard; Protein; 1430 AA.  
XX  
AC AAU98043;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE S. mutans glucosyltransferase GTFD mutant N471D.  
XX  
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
KW coating composition; glucan; starch; latex; thermoplastic molecule;  
KW amyloplast; vacuole; paper manufacture; mutant; mutain.  
XX  
OS Streptococcus mutans.  
QS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 471  
FT /note= "Wild-type Asn substituted by Asp"  
XX  
PN US2002031826-A1.  
XX  
PD 14-MAR-2002.  
XX  
PF 19-DEC-2000; 2000US-0740274.  
XX  
PR 11-DEC-1998; 98US-0210361.  
PR 07-JUN-1995; 95US-0478704.  
PR 07-JUN-1995; 95US-0482711.  
PR 07-JUN-1995; 95US-0485243.  
PR 16-JAN-1998; 98US-0007999.  
PR 20-JAN-1998; 98US-0008172.  
XX  
PA (NICH/) NICHOLS S E.  
XX  
PI Nichols SE;  
XX







QY 1718 LVQT-----DNGTYFDNYGK-----MVRNOTINAGAMIYTLDEN 1752  
 Db 1161 FLENADGSKNYFCHLGNRYNGYISFNDKSKRYFDASGVNAVGLKTINGTQYF--DQD 1218  
 QY 1753 GKLRKASY 1760  
 Db 1219 GYQVKGAW 1226

RESULT 9  
 AAU98029  
 ID AAU98029 standard; Protein; 1430 AA.  
 XX AC  
 XX AAU98029;  
 DT 27-AUG-2002 (first entry)  
 XX DE  
 XX S. mutans glucosyltransferase GTFD.  
 XX KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX OS Streptococcus mutans.  
 XX PN US2002031826-A1.  
 XX PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-0740274.  
 XX PR 11-DEC-1998; 98US-0210361.  
 PR 07-JUN-1995; 95US-0478704.  
 PR 07-JUN-1995; 95US-0482711.  
 PR 16-JAN-1998; 95US-0485243.  
 PR 16-JAN-1998; 98US-0007999.  
 PR 16-JAN-1998; 98US-0008172.  
 PR 20-JAN-1998; 98US-0009620.  
 XX FA (NITCH/) NICHOLS S E.  
 PI Nichols SE;  
 XX WPI; 2002-414332/44.  
 DR N-PSDB; ABK52940.  
 XX

Glucosyltransferase B or D protein useful for producing a glucan useful  
 as substitutes for and additions to modified starch and latexes in  
 paper manufacture, comprises mutations in specific positions -  
 PS Disclosure; Page 38-42; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase  
 CC (GTF) B polypeptide having changes at position from 1448V, D457N,  
 CC D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K,  
 CC D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K790/K1014T,  
 CC Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having  
 CC changes at positions from T589D, T589E, N471D, N471D/T589D, and  
 CC N471D/T589E. Also included are a glucan produced by the GTF mutant,  
 CC an isolated polynucleotide which encodes P1 or P2, or its complementary  
 CC polynucleotide, a ribonucleic acid sequence encoding the GTF mutant,  
 CC an expression cassette comprising the polynucleotide operably linked to a  
 CC promoter, a vector comprising the expression cassette, host cell  
 CC introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex, where  
 CC thermoplastic molecule or their combinations of glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step.  
 CC The present sequence represents GTFD.  
 XX

SQ Sequence 1430 AA;  
 Query Match 26.0%; Score 2446; DB 23; Length 1430;  
 Best Local Similarity 46.7%; Pred. No. 5.8e-128;  
 Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

QY 738 LOTINGQYYIDPTTGQPRKNFLQSGNNWYFDSDTGVGTNALEQFAKGVSSNEQYR 797  
 Db 174 VKOIDGKYIIG--SDGQPRKNFALTNNKVLFDKNTGALTDTSQYQFGTKLNDYT 232  
 QY 798 NGNAYSYDKSIENVNGYLTADTWYRKOILKDGSTTWTDSKETDMRPLMYWNPNTLTQ 857  
 Db 233 PHNOIVNFENTSLETDVNTVADSWYRPRDKLNGKTWTATASSESDLRPLLLMSWPKQKQ 292  
 QY 858 AYVLNMYKQHG-----NLLPSALPFENADAPALNHYSEIYQOQNEKRISETGNTDWR 912  
 Db 293 IAYLNYMNOOGLGTGEN-----YTADSSQESLNLAAQTQVVKIETKISQTOQTOWLR 344  
 QY 913 TLMHDEVTNNPMWKNKDSNVNFSGIQ--FQGGFLKYENSLTPYANSYRLLGRMPIN-- 968  
 Db 345 DIINSEVKIQPNWNSQTESDSAGEKDLQGGALLYNSDKTAYANSYRLLNRTPTSQT 404  
 QY 969 -----IKDOTYRQOEFLANDIDNSNPVQAEQLNWLNYLLNFGTTTANDQANEDSV 1023  
 Db 405 GKPKYFEDNSSGGYDFELLANDIDNSNPVQAEQLNWLNYLLNFGTSIVANDPANFQGV 464  
 QY 1024 DAPDNIDADLMNTAQDYFNAAYGMD--SDAVSNKKHINILEDWNHADPEYFNKIGNPQ 1082  
 Db 465 DAVDYNADLLQIASDYLAHYGVDSKSEKNAIHLSEILAWSNDNDPQYKDKGAOLPID 524  
 QY 1083 DTKNSLNHGLS-----DATN----RWGLDAIVHQSLADRENNSTENVIPYSFVRAHD 1133  
 Db 525 NKLRLLLYALTPLKEDASNKNEIRSGLEPVTNLSNRRSAEGKNSERMANIYFIRAHD 584  
 QY 1134 NNSQDQIQONAIR--DVTGK--DYHTFTFEDQKIGIDAVIQDNSTVKKYKLYNIPASVAIL 1191  
 Db 585 SEVQTVIAKTIKAQINPKTDGLTFTLDELKQAPKIYNEQDMRQAKKKYQTQSNIPTAYAL 644  
 QY 1192 TNKDDTPRVYIGDLYDGGQYMEHQIRYDITLNLKSRVKYVAGQSQWTSVSGVGN-- 1250  
 Db 645 SNKDSITRILYGYMDYSDGGQYATKSPYDAIDITLKARIKAAAGGQDKKITYVEGDKSH 704  
 QY 1251 -----ILTSYRVYKGAEMATDTGTDETRTOGIVGVVSNTPNLKLGVDNKKVVLHMGAAH 1303  
 Db 705 MDWDYTGVLTSYRVYKGAENATDQGEAKTKQGAIVTSNNPSLKLNDQDKVIVNMGAAH 764  
 QY 1304 KNOQYRAAVLTITDGVINYTSDDGAP--VAMTDENGDLXLSHNLVNVNGKEADTAVQOY 1361  
 Db 765 KQGEYRPLLTITKDGLTSTDAAKSLYRKTKNDKGELVFDASD-----IQGY 812  
 QY 1362 ANPDVSGYLAVVPVGVASQNDQARTAPSTEKNSGNSAYTNAAFDSNVIFAESFNYVTP 1421  
 Db 813 LNPQVSGYLAVVPVGVASQNDQVRVAASKNANAIGOVYESSSALDSQLIYEGFSNQDFV 872  
 QY 1422 TKSESRANRYAQNADFFASLGFTSEMAPQYNSSKDRFTFLDSTDINGYAFYDRLGHS 1481  
 Db 873 TKSDSYTNKIAQNQVLFKSWGVTSEMAPQYVSSDGGSLDSTIQNGYAFEDRYDLAMS 932



Db	443	PVQABQLNWLHFLNFGNIYDNDPANDPSIRVDVNDVNDDALLQIAGDYLKAAKGIIHK	502
Qy	1049	SDAVSNKHINILEDNNHADPEYFNKIGNPQLTMDDTIKNSLNHGLSDATN-RWGLDAIVH	1107
Db	503	NDKAAANDHLSILEANSYNDTPYLDHGDGNMNLNRRLSLSLYLAKPLNQRSGMNPLOT	562
Qy	1108	QSLADRENNSTENVVYPNKSFVRAHDNNSQOIONAIR- ---DVTCKDHYHTTFEDEOK	1162
Db	563	NSLVNRTDDNATAAPVSYSFIRAHDSVQDLIRNIIITEINPNVG- ---YSFTTEIKK	619
Qy	1163	GIDAYTODQNSTVKKYNLYNIPASVAILLTNKDIPRVYVYGDLYTDGGVMEHOTRYDIT	1222
Db	620	AFEYINKDULLATEKKYTHYNTALSVALLTNKSVPVRYVYGDMTDGGQYMAHKTINYE	679
Qy	1223	LTWLLSRVYKVVAGGOSMOTMSVGCNNLILSVRYGKGAAMTATGTDETRTQIGVYVVS	1282
Db	680	IFELLKARKIKYVSGGOMRNOQV- GNSELIISVRIGKALAKATIDGRTTRISGVAVIEG	738
Qy	1283	NTPNLKLGVNDKVVLHMGAAHNQOYRAVLTTTGDGVINYTSDQAP- -VAMTDENGDLY	1340
Db	739	NNPSLRLKASDRVVNMGAAHNQAYRPLLTTDNGIKAYHSDGEAAGLVRYTNDRGELI	798
Qy	1341	LSSHNLVWCKEADRAVOGYANPDVSGVLAWVPVGASDNDQARTAPSTEKNSGNSAYR	1400
Db	799	FTA- ---AD- ---IKGYANPOVSGVLGVWVPVGAADQDVRVAASTAPSTDGKSVH	846
Qy	1401	TNAAFDSNVIFEAFSNFYPTTKESERANVRIAQNADFFASLGTFTFEMAPQYNSKDRT	1460
Db	847	QNAALDSRVMEFGFSFQAFATKKEEYTNVVIKKNVDAFEWGYTDFEMAPQYVSSYDGS	906
Qy	1461	FLDSTDNGVAFTRDYRDLGMSBPKNYGTDEDLRNIAIQALHRAQLQVMAWVDPDQIYNLPG	1520
Db	907	FLOSIVONGVAFTRDYLGISPKNYGTADLLVYKAIKALHSGIKYMAWVDPDQYALPE	966
Qy	1521	KEYATVTRVDDRGNWKKDAIINNLYVYN-TIGGEYOKTKYGGAFDLKLQKLYPEITTKK	1579
Db	967	KEYVTRATRVKGYTPVAGSQIENTLYVVDGKSSGQDQAKYGGAFLEELQAKYPELEFARK	1026
Qy	1580	QVSTGVAIDPSOKITFWSAKYFNGLNLRHSGVYVKADGGQYNLGTTTKQFPQIOLIT	1639
Db	1027	QISTGVPMDSVKIKOWSAKYFNGLNLRGACVYVKQATNTVFSLVSDNTFLPKSLV-	1085
Qy	1640	EKKQNGEFGYKG- ---NDG- NYFYDLAGNMVKNFTIEDSVG- NWIFFDQGGKVENKHUFVD	1695
Db	1086	NPNHGTSSTVGLVFDCKGYVYYSTSGNOAKNAFI- -SLGNWYVFDNNGYVYVGAOSIN	1143
Qy	1696	VDSYGEKGTYFFLKNGVSPRGGIYQV- ---DNGNY- ---YFDNVGK	1733
Db	1144	-----GANYFYLSNGIQLRNALYDNGKNVLSYDNGRRVNGYLFEGQWRFFQN-GI	1196
Qy	1734	MVRNOTINAGAMITYITLDENGKLKASYNDSAEYPTSTDVGKM	1775
Db	1197	MAVGLTRVHGAVOY-FDASG- -----FOAKGQFITIAD-GKL	1230

RESULT 11

RESULT I:  
AA179288

AAU 75200  
TD AAU

XX  
ID  
XXAC  
AAU'

XX

DT 13-7

XX

DE Stro

XX

KW Ant:

KW ant:

XX

SO  
str

XX  
DNI  
TDCXX  
FN  
JFZ

XX 16-

NOTED

04-OCT-2000; 2000JP-0304889.  
04-OCT-2000; 2000JP-0304889.  
(UYN1-) UNIV NIPPON.  
WPI; 2002-448101/48.  
Anti-carries agent composed of a monoclonal antibody against an inhibitory enzyme against water insoluble glucan synthetase of glucosyl transferase-B (GTF-B) of Streptococcus mutans -  
Disclosure; Page 22-25; 28pp; Japanese.  
The invention relates to a monoclonal antibody against dental caries and an anti-carries agent composed of a monoclonal antibody produced by Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566) or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having inhibitive activity against water insoluble glucan synthetase of glucosyl transferase-B. The monoclonal antibody specifically inhibits water insoluble glucan synthetase of Streptococcus mutans produced glucosyl transferase-B and is used in the immunotherapy of dental caries. This sequence represents a Streptococcus mutans monoclonal antibody-related protein.  
Sequence 1375 AA;  
SQ

Query Match 24.9%; Score 2347.5; DB 23; Length 1375;  
Best Local Similarity 41.3%; Pred.No. 1.8e-122;

Qy	549	SAVSGFDTT	IKLTNDQYALNGLOVLRF	SKAADG--NPSGDNTVTVDQFSKNYATTCGN	606
Db	29	TLISGLSVKADSDRQAVTESQAASLVTTSEAAKEITLTDSTATSQPPAIVTDN	88		

QY 607 FDYKVNQVFEFGHATNQSNKDKDSQWIIVLVNGKEVKRQLVNDTKEGAAGFNRDVY 656

09 VS ----- LINQ3  
DU 667 KVNPAIENSSMSGFOGII TLPVTYKNNVOLVHRESNDVKTEGNV-----DFW 716

Db 108 KPTTSEQAKTNSDKIITTSKAVNR-----LTATGKRVPPANNNTAHPKTVT 154

QY 717 SELPVRKDSFGKNGPLKQ-----FGLQTLNGQQYIDPFTTGQPKRNE LL 701

QY 762 QSGNNWYFDSDTGVTNALEIQFAKGTVSSNE---QYRNGNAAYSYDDKSIENNVGILT 818

Db 210 NINGKTFEDETGALSNTLPSK--KGNITNDNTNSPAQYNQYVSTDVANFEHVDHYLT 267

```
Db      |::||| |||| ||| | | ||| ||| : : :  
268 AESWYRKYILDKGKTWTQSTEDFRPLMTWWPDQEIQRVVYMNAQLGIHQ-----Y 323
```

QY 879 NADAPAEINHSEIVQONIEKRISSETGNTDNLRLMHDFVTNNPMWNKDSNVNFSIQ 938

04 939 FQGQFLKYNSDTPYANSYRLLGRMPIN-----IKDQTYRGQFELLANDIDNSN 989

384 -KGALLYSNSKLT SQANSYRIILNRTPTNQTKKDPYATADRTIGGYEFLANDVDNSN 442

QY 990 PVQAEQUNWLY I LUNFG I ITANNDQANE DSVRVDAFONLADADMNIAQITNKGALOM D 10490  
443 BVVQAEQI MWI HFETANFEGITVANIIPDANEDSTRVDVNVNDADLLOIAGDYLKAAQGIHK 502

QY 1049 SDAVSNKHINILEDOWNHADPEYFNKIGNPQLTMDTIKNSLNHGSLDATN-RWGIDAIVH 1107

Db 503 NDKRANHL5ILEAWSYNDTPYLHDDGDNMNDNRRLRSLLSLAKPLNQSRGNPLIT 562

Db 563 NSLVNRDNDNAETAAPSYSPFIRAHDSVQDLIRNIIRTEINPNVWG---YSETTEBIKK 619  
Qy 1163 GIDAYIODNSTVKYKLYNIPASVAILLTKNDIPRVYXGDLTDCGQYMEHCOTRYDYT 1222  
Db 620 AFEIYKDLATEKRYTHYNTALSYALLLNKSSVPRVYGDFTDGGQYMAHKTINVEA 679  
Qy 1223 LTNLLKSKYVAGQSQMOTMSVGGNNILTSVAYGRKAMTADTGTDETRTGGIGYVVS 1282  
Db 680 IETLLKARIKIVSGQAMRNQOV--GNSIITSVRYGKALKATDGTDRTRTSGVAVIEG 738  
Qy 1283 NTPNKLGVNDKVLHGAHKNQYRAAVLTITTDGVINTVSDGAP--VAMTDENGDLX 1340  
Db 739 NPFSLRLKASDRVVVNGAAGKNGAYRPLLLTDDNGIKAYHSDQEAAGLVYINDRELI 798  
Qy 1341 LSSHNLVYNGKEADTAVQGYANPDVSGYLAVWVPGASDNQDARTAPSTEKNSGSAIR 1400  
Db 799 FTA-----AD--IKGYANQVSGYLGWVVPVGAADQDVRVAASTPSTDGKSVH 846  
Qy 1401 TNAAFDSNVIFEARNSFYTTKESERANVRIAGNADFFASLGFTSEMAPOYNSSKDR 1460  
Db 847 QNAALDSRVMEFGFSNFOAFATKKEEYTNVIAKNVDKFAEWGVTDFEMAPQYVSSYDGS 906  
Qy 1461 FLDSTIDNGYAFTDRYDLGMSSEPNKYGTDEDLRNAIQALHKAGIQVMADWVPDQIYNLPG 1520  
Db 907 FLDVSIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGIKVMADWVPDQYALPE 966  
Qy 1521 KEVATVTRVDRGNVWKDAIINNLYVN--TIGGEYQKYGGAFLDKLQKLPDPEITTK 1579  
Db 967 KEVATVTRVDRGNVWKDAIINNLYVN--TIGGEYQKYGGAFLDKLQKLPDPEITTK 1579  
Qy 1580 QVSTGVAIDPSOKITENSAKYFNGTNIIHRSGYVVLKADGGQYVNLGTTTKQFLPIOLTG 1639  
Db 1027 QISTGVMPDPSVKIKQNSAKYFNGTNIIHRSGYVVLKADGGQYVNLGTTTKQFLPIOLTG 1639  
Qy 1640 EKKQNEGFVKG--NDG--NYFYDLAGMVKNTFIEDSVG--NWYFFDODGKMWENKHFVD 1695  
Db 1086 NPNHGTSSSVTGLVFDGKGYVYVYSTSGNAKNAPF--SLGNWYTFDNGYVMTGAQSIN 1143  
Qy 1696 VDSYGEKGYTFELKNGYVFRGLVQT-----DNGTY-----YEDNGK 1733  
Db 1144 -----GANYFVLSNGIOLRNAIYDNGKVLVSYGNDGRYENGYYLFGQWRYFQN-GI 1196  
Qy 1734 MYRNQITINAGAMITLDENGLIKASYNSDAEYPTSTDVGKM 1775  
Db 1197 MAVGLTRVHGAVQY-FDASG-----FOAKGQFITAD-GKL 1230

RESULT 12  
AAU79284  
XX ID AAU79284 standard; Protein; 1476 AA.  
XX AC AAU79284;  
XX DT 13-AUG-2002 (first entry)  
XX DE Streptococcus mutans monoclonal antibody-related protein #1.  
XX KW Antibody; dental caries; water insoluble glucan synthetase;  
XX OS anti-carries; glucosyl transferase-B; immunotherapy.  
XX OS Streptococcus mutans.  
XX PN JP2002114709-A.  
XX PD 16-APR-2002.  
XX PF 04-OCT-2000; 2000JP-0304889.  
XX PR 04-OCT-2000; 2000JP-0304889.  
XX FA (UNIV-) UNIV NIPPON.  
XX

DR WPI; 2002-448101/48.  
XX Anti-carries agent composed of a monoclonal antibody against an  
PT inhibitory enzyme against water insoluble glucan synthetase of glucosyl  
PT transferase-B (GTF-B) of Streptococcus mutans  
XX  
PS Claim 3; Page 13-16; 28pp; Japanese.  
XX  
CC The invention relates to a monoclonal antibody against dental caries and  
CC an anti-carries agent composed of a monoclonal antibody produced by  
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
CC transferase-B. The monoclonal antibody specifically inhibits water  
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
CC transferase-B and is used in the immunotherapy of dental caries. This  
CC sequence represents a Streptococcus mutans monoclonal antibody-related  
CC protein.  
XX  
SQ Sequence 1476 AA;  
Query Match 24.7%; Score 2330.5; DB 23; Length 1476;  
Best Local Similarity 42.8%; Pred. No. 1.7e-121;  
Matches 522; Conservative 185; Mismatches 379; Indels 135; Gaps 35;  
Qy 637 IYLVNGKEVKRQLVND-----TKEGAAGFNNDVYKVPAINSSM 677  
Db 35 LVKADSNESKQSINSDNTSVTANDESNVITEATSKQEAASSQTN--HIVTSSSSVS 92  
Qy 678 SGFQGITLPTVTVKNNVQLVHRFSNDVKTGEGNYVDFWSEL-----MPVKD 724  
Db 93 VNPKEVNSPYTV-GETASNGEKLQNTTT-----VDKTSEAAANNISKOTTEADTVID 146  
Qy 725 SFQKGNGL--KQFGLQTNGQYIIDPTGQPRKNFLQSGNNWIYFD----SDTGVG 777  
Db 147 DSNAAQLQLEKLPNYKEIDGKYIYD--NKGKVTNFTLIADGKILHFDGTGAYDTSID 205  
Qy 778 TNAELQFAKGTYSSENEQYRNGNAAYSYDDKSIERNVGYLTADTWYRPQILKDGTTWD 837  
Db 206 TVNKDIV---TTRSN-LYKKYNOVYDRSAQSFHVDHYLTAEWSYRPYILKDKGTWTQ 260  
Qy 838 SKETDRPILMWPNWPTLQYLYNTKMHGHNLLSALFFFNADADPAELNHYSELVQON 897  
Db 261 STEKDFRPLMTWMPDOETQRYVYNNQAQLGINKT----YDTSNQLNIAAATIAQAK 316  
Qy 898 IEKRISETGNTDMLRTHMDFVTNNPMWKNKSDENVNFSGIQGGFLKYEN--SDLTPTYAN 956  
Db 317 IEAKITTLKNTDLROTISAFVKTQSAWNSDSEK-PFDD-HLQNGAVLYDNEGKLTPTAN 374  
Qy 957 SDYRLLGRPIN-----IKDQYRQGEFLANDIDNSNPVVOAEOLNWLIIYLLNFG 1007  
Db 375 SNYRIILNRTPTNOTKKDPYATADNTIGGYEFLANDVDNSNPVVOAEOLNWLIIYLLNFG 434  
Qy 1008 TITANNDAQNFDSVRVDAPDNIDADLMNTAQDYFNAAYGM--DSDAVSNKHIILEDWNAH 1066  
Db 435 NIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAKGHKKDKAANDHLSLEAWSN 494  
Qy 1067 DPYFNKKNIPQLTMDDTKNSLNHGLSDATN--RWGLDAIVHQSADRENNTENVTN 1125  
Db 495 DTPYLDHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMNPFLITNSLNRFTDQNAETAAVPS 554  
Qy 1126 YSFVRAHDNNSODQIONAIR----DVTGKYHTFTFEDEQKIDAYIODNSTVKKYVL 1180  
Db 555 YSFIRAHDSVQDLIRDIKAEINPNVWG---YSFMEBIKKAFFIYNKOLLATEKKYTH 611  
Qy 1181 YNIPASYAILLTKNDIPRVYXGDLTDCGQYMEHCOTRYDYTTLNLLKSRVYVAGGQSM 1240  
Db 612 YNTALSYALLLNKSSVPRVYGDFTDGGQYMAHKTINVEAETLLKARIKIVSGQAM 671  
Qy 1241 QTHSVGNNNLTSTVRYGKAMTATDTGTDETRTGGIGYVVSNTPNLKGVDKVLHMG 1300  
Db 672 RNOQV--GNSIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPSLRUKASDRVVWNG 730

QY	1301	AAHKNOQYRAAVLTTDGVINTSQDAP--VAMTDENGDIYLSHKLNVNGKEADTAV	1355
Db	731	AAHKNOAYRPLLLTDNGIKAYHSQEAAGVRYTNDRGELIFTA-----AD--I	778
QY	1359	QGVANPDVSGYLAVWPVPGASNDQARTAPSTEKNSGNSAVRTNAAFDSNVIFEAFSNFV	1418
Db	779	QGVANPDVSGYLAVWPVPGAAADQDVRVAASTAPSTDGKSVHQNAALDSRVWFEFSNFQ	838
QY	1419	YPTFKESERANVRIQONADFFASLGFTSFENAPQYNSKQRTFLDSTDIDNGYAFTRDYDL	1478
Db	839	AFATKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVYSSDGSFLDSVIQNGYAFTRDYDL	898
QY	1479	GMSEPNKYGTDEDLRAIQALHAKGLQVMADWVPDQIYNLPKGKEVATVTRYDDDRGNWKD	1538
Db	899	GISKPNKYGTADLLYKAIKALHSKIKVMADWVPDQMTAFPEKEVYATRVYDKYGTVPAG	958
QY	1539	AIINNLNLYVN--TIGGEYQKKYGGAFLDKLOKLYPEITFTKKQVSTGVAIDPSOKITEMS	1597
Db	959	SOIKNTLYVVDGKSSGQDOQAQYGGAFLELQAKYPELFARKQISTGVPMDEPSVKIKQWS	1018
QY	1598	AKYFNGTNIILHRGSGVYLKADG--GOYYNLGITTK--QELPIQLTGEKKOGNSEFVKGNNGN	1655
Db	1019	AKYFNGTNIILGRGAGVYLDQDQATNYFNISDNKEINFLPKTLNLDQSO--VGFSYDQKG-	1075
QY	1656	YFYPDLAGNNKNTFIEDSGVNWYFFDQDGKMKVENKHFVDVDSYGEKGTYPFLKNGVSFR	1715
Db	1076	YVYISTSGYQAKNTFISEG--DKWYFDNNGYVMTGAQSI-----GVNYTFLSNGLQLR	1128
QY	1716	GGLVQTDNGTY-YFDNYGKVMRN-----QTNAGAM-----IYTLDENGK	1754
Db	1129	DAILKNEGDYIAYGNDGRYENGYQFMGVYRHFNGEMSVGLTVIDGQVQYFDEMG-	1187
QY	1755	LIKASYNSDAEYPTSTDVGKM	1775
Db	1188	----YQAKGKEVTTAD-GKI	1202
RESULT	13		
AAU98033	ID	AAU98033 standard; Protein; 1475 AA.	
XX	AC	AAU98033;	
XX	DT	27-AUG-2002 (first entry)	
XX	XX	S. mutans glucosyltransferase GTFB mutant K1014T.	
DE	XX	Glucosyltransferase; GTFB; transgenic plant; paper sizing;	
KW	KW	coating composition; glucan; starch; latex; thermoplastic molecule;	
KW	KW	anyploplast; vacuole; paper manufacture; mutant; muten.	
XX	OS	Streptococcus mutans.	
OS	XX	Synthetic.	
XX	Key	Location/Qualifiers	
FH	FT	Misc-difference 1014	
FT	FT	/note= "wild-type Lys substituted by Thr"	
XX	PN	US2002031826-A1.	
XX	PD	14-MAR-2002.	
XX	PF	19-DEC-2000; 2000US-0740274.	
XX	XX	11-DEC-1998; 98US-0210361.	
PR	PR	07-JUN-1995; 95US-0478704.	
PR	PR	07-JUN-1995; 95US-0482711.	
PR	PR	07-JUN-1995; 95US-0485243.	
PR	PR	16-JAN-1998; 98US-0007599.	
PR	PR	16-JAN-1998; 98US-0008172.	
XX	XX	20-JAN-1998; 98US-0009620.	
XX	XX	(NICH/) NICHOLS S E.	





Db	836	FOAFATKKEEYTNWIAKNVDFKFAEAGVYDFEFMAPOYVSSSTDGSLDSVIONGFAFTDRY	895
Qy	1477	DLQMSPEPKYGTDEDLRNATQALHKAGLOMADWPDQGTINLPGRKEVATVTRVDDRGNVW	1536
Db	896	DLGISKPNKYGTADDLVKAIKALHLSKGIKVMADWPDQMYAFPEKEVVTATRVDKYGTPTV	955
Qy	1537	KDAIINNLLVNV-TTGGGEYKKYGGAGFLDKLQKLYPEIFTKKQVSTGVAIDPSQKITE	1595
Db	956	AGSQIKNTLVVDPKSSGKDQQAQYGGAGFLELQAKYPELFARKQISTGVPMDPSPVKIKQ	1015
Qy	1596	WSAKYFNGTNIILHRGSGYVLKADG-GQYYNLTGTTK-QFLPIQLGKQKQNGEFGVKGND	1653
Db	1016	WSAKYFNGTNIILRGAGYVLKQATWTYENISDNKEINFLPKTLNQDSQ--VGFSYDGK	1073
Qy	1654	GNIFYDLAAGWYKNTFIEDSVGNWTFYFDDQSKMYENKHFVDVDYSYGEKGYFFFLKNGVS	1713
Db	1074	G-YVYVSTSGYQAKNTFISEG-DKWFYFDNNGYMTGAQSIN-----GVNYEFLNSGLQ	1125
Qy	1714	FRGGLGDTGNGTY-YFDNTYGMVNR-----QTINAGAN-----IYTLDEN	1752
Db	1126	LRDAILKNEDGTAYYCNDRGRREYGYQFMGSQVHRHFNNGEMSVGLTVIDQGVYFDDEM	1185
Qy	1753	GKLIKASYNSDAEYPTSTDVGKM	1775
Db	1186	G-----YQAKGRFVTAD-GKI	1201
RESULT 15			
AAU98030			
ID	AAU98030 standard; Protein; 1475 AA.		
XX			
AC	AAU98030;		
XX			
DT	27-AUG-2002 (first entry)		
XX			
DE	S. mutans glucosyltransferase GTFB mutant T448V.		
XX			
KW	Glucosyltransferase; GTFB; transgenic plant; paper sizing;		
KX	coating composition; glucan; starch; latex; thermoplastic molecule;		
KW	anyploast; vacuole; paper manufacture; mutant; mutain.		
XX			
OS	Streptococcus mutans.		
OS	Synthetic.		
XX			
XX			
Key	Location/Qualifiers		
FH	448		
FT	Misc-difference		
FT	/note= "wild-type Ile substituted by val"		
XX			
PN	US2002031826-A1.		
XX			
PD	14-MAR-2002.		
XX			
PF	19-DEC-2000; 2000US-0740274.		
XX			
XX			
PR	11-DEC-1998; 98US-0210361.		
PR	07-JUN-1995; 95US-0478704.		
PR	07-JUN-1995; 95US-0482711.		
PR	07-JUN-1995; 95US-0485243.		
PR	16-JAN-1998; 98US-0007999.		
PR	16-JAN-1998; 98US-0008172.		
PR	20-JAN-1998; 98US-0009620.		
XX			
PA	(NICH/) NICHOLS S E.		
XX			
PI	Nichols SE;		
XX			
DR	WPI; 2002-414332/44.		
XX			
XX			
PT	Glucosyltransferase B or D protein useful for producing a glucan useful		
PT	as substitutes for and additions to modified starch and latexes in		
PT	paper manufacture, comprises mutations in specific positions -		
XX			

Claim 36; Page -; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36.

Sequence 1475 AA;

Query Match 24.2%; Score 2277; DB 23; Length 1475;  
Best Local Similarity 42.4%; Pred. No. 1.7e-118;  
Matches 519; Conservative 182; Mismatches 382; Indels 140; Gaps 37;

Qy 637 IVLVNGKEVKQLVND-----TKEGAAGFNNDVYKYNPAIENSSM 677  
Db 35 LVKADSNESKQISNDNSNTSYVTANEESNVITEATSKQEAASQTN--HVTTSSTSTV 92  
Qy 678 SGFGGIIPLVTVRNENVLVRFSNDVKTEGNYVDFWSEL-----MPVKD 724  
Db 93 VNPKEVYSNPYV-GETASNGEKLQNTT-----VKTSEAAANNISKQTTEADTVID 146  
Qy 725 SFQKNGPL--KQGLQTINGQQYIDPTTGQPKRKNFLOSNNWYFD-----SDTGVG 777  
Db 147 DSNAAQLQLEKLPNVKEIDKYYIYD--NNGKVRTNETLIADGKILHFDGTGAYDTSID 205  
Qy 778 TNALQLQAKGTSSNQQYRNGNAAYSYDVKSIENVNGYLTADTWYRPPKQILKDGTTWTD 837  
Db 206 TVNKDIV-----TTRSN-LYKKNYGVDRSAQSFPHVYLTAEWSYRPPKYLKDGKTTW 260  
Qy 838 SKETDMRPLMWWPNTLTQAYLYNKMKGHGNLLPSALPFNFADAPAEALNHYSEIVQQN 897  
Db 261 STENKDFRLMTWPDQETQYQYVYNNQAQGINKT---YDFTSNQLQNLNIAAATIQAK 316  
Qy 898 IEKRISGTNDWLRLTLDHDPVTNNPMWNKDSNNVNSGTFQCGFLKYN--SDLTTPYAN 956  
Db 317 IEAKITLKNLQTLRQITAFVKYQCSAWNSDSEK-PFDD-HLQNGAVLYDNEGKLTTPYAN 374  
Qy 957 SDYRLGKRPIN-----IKQGYRGQEFLLANDIDNSNPVYQAEQLNWLNYLLNFG 1007

Db 375 SYNRLNRTPTTQTKKKDPRTADNTTIGGYEFLLANDVDNSNPVYQAEQLNWLHFMNFG 434  
Qy 1008 TITANNQANFDSVRVADPDNTDADLMTIAQDYFNAAYGM-DSDAVSNKHNILHELDWHA 1066  
Db 435 NIYANDPDANFDSVRVADVDVNDVADLLQTAGDYLAKAAGIHKNDKAANDHLSILEAWSN 494  
Qy 1067 DPEYFNKIGNPOLTWDOTTKNSLNHGLSDATN-RGLDAIVHQSLADRENNSTENVVPIPN 1125  
Db 495 DTPYLHDDGDNMINMDNKURLSLSLAKPLNQRSGMPLITNSLWNRDDEAETAAPVS 554  
Qy 1126 YGFRVRAHNSDQIQONAIR-----DVTGKDYHTTFFDEQKIDAYIQDQNSTYKKNYL 1180  
Db 555 YGFIARHDEVDQIADIADIKAEINPVVG---YSTMERIKKAFETYNKDILLATEKETH 611  
Qy 1181 YNIPASVAILLNKOTIPRVYVGDLYTDGQYMEHQTRVYDILTLLKSRVYVAGGSM 1240  
Db 612 YNTALSYALLNKSVPRVYVGDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAM 671  
Qy 1241 QTMVSGNNILTSVRYGKAMTADTGTDETGTIGIGYVVVSTPNLKLGVNDKYVLHMG 1300  
Db 672 RNQCV-ENSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPFLKASDRVVVNG 730  
Qy 1301 AAHKNOQYRAAVLTITDGVINTSDGAP--VAMTDENGDLVLSHNLVNGKEADTAV 1358  
Db 731 AAHKNGAYRPLLTDDNGIKAYHSDQEAAGLVRYNDRGELIFTA-----AD--I 778  
Qy 1359 QGYANDVSGYLVAVVYPVGAS--DNODARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSN 1416  
Db 779 KGYANQVSGYLVGVVYVGAALIKFALRLA---RPHQQMASVHQVNAALDSRVMEGFEN 835  
Qy 1417 FVYPTPKESERANVRTAQNADFASLGFTSFEMAPQYNSKSDRTFLDSDIDNGYAFDTRY 1476  
Db 836 FQAFATKKEEYTNVJAKNVDFAEWGVDFEMAPQYVSSDTGSLDSVIQNGYAFDTRY 895  
Qy 1477 DLGMSFPNYGTDEDLRNAIQALHAKGLQVADWVDPDQIYNLPGREVATVTRVDDRGNYV 1536  
Db 896 DLGISKPNKYGTADLLVKAIKALHSGIKVMADWVDPQMYAPPEKEVVTATRVKYGTPV 955  
Qy 1537 KDAIINNLYVYN-TIGGGEYQKYGGAFLDKLOKLYPEIFTKKQVSTQVADPQOKITE 1595  
Db 956 AGSQIKNTLYVVDKSSGKDDQAKYGGAFLEELQAKYPELFARKQISTGVPMDDPSVKIKQ 1015  
Qy 1596 WSAKYFNGTNILHRSGYVLKADG-QQYVNLGTTTK-QFLPIQLTGEKKQNGEYFKGND 1653  
Db 1016 WSAKYFNGTNILGRGAGYVLKQATNTYFNISDNKEINFLPKTLLNODSQ--VGFSYDGK 1073  
Qy 1654 GNYFYDLAGNMVKNFTFIEDSGVGNWYFFDQDKWVGNKHFDVDSYGEKGTFFFLKNGVS 1713  
Db 1074 G-YVYVYSTSGYQAKNTFISEG-DKWWYFDNNGYMYVTGAQSN-----GVNYFYLSNGLQ 1125  
Qy 1714 FRGGLVQTDNGTY-YPDNYGKVVNR-----QTINAGAM-----IYTLDEN 1752  
Db 1126 LRDAILKNEDGIYAYVYGNDRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQVQYFDEM 1185  
Qy 1753 GKLIKASYNSDAEPYFTSDVGMK 1775  
Db 1186 G-----YQAKGKEVTTAD-GKI 1201

Search completed: April 23, 2003, 15:07:41

Job time : 123 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 15:05:37 ; Search time 21 Seconds

(without alignments)  
2495.344 Million cell updates/sec

Title: US-09-995-749A-2

Perfect score: 9424

Sequence: 1 MEIKKFKLYKSGKQWTRAA.....SDAEYPTSDVGKMLDQNK1 1781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_RA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2446	26.0	1430	3	US-09-008-172-2
2	2446	26.0	1430	4	US-09-210-361-6
3	2446	26.0	1430	4	US-09-740-274-6
4	2347.5	24.9	1375	4	US-09-210-361-4
5	2347.5	24.9	1375	4	US-09-740-274-4
6	2276	24.2	1475	3	US-09-007-999-2
7	2276	24.2	1475	4	US-09-210-361-2
8	2276	24.2	1475	4	US-09-740-274-2
9	2190.5	23.2	1577	2	US-08-793-824-2
10	327.5	3.5	2411	4	US-09-268-347-36
11	326.5	3.5	349	3	US-09-009-620-2
12	321.5	3.4	2353	4	US-09-377-155-33
13	321.5	3.4	2353	4	US-08-913-942-4
14	321.5	3.4	2353	4	US-09-669-374-33
15	318.5	3.4	2314	4	US-09-268-347-49
16	317	3.4	2354	4	US-09-268-347-47
17	303	3.2	10182	4	US-09-134-001C-3159
18	292	3.1	3696	4	US-09-134-001C-5080
19	284	3.0	1912	1	US-08-409-995-4
20	284	3.0	1912	3	US-08-685-467-4
21	278	2.9	1335	4	US-09-134-001C-3716
22	277.5	2.9	1323	4	US-08-988-685A-10
23	271.5	2.9	2048	4	US-09-268-347-48
24	261	2.8	1833	4	US-08-621-944A-4
25	261	2.8	1833	4	US-08-945-567D-4
26	261	2.8	1992	4	US-08-621-944A-3
27	261	2.8	1992	4	US-08-945-567D-3

```

28 258.5 2.7 2137 4 US-09-134-001C-4463 Sequence 4463, Ap
29 256.5 2.7 2710 1 US-08-480-604A-6 Sequence 6, Appli
30 256.5 2.7 2710 2 US-08-405-496A-6 Sequence 6, Appli
31 256.5 2.7 2710 4 US-08-915-136-6 Sequence 6, Appli
32 256.5 2.7 2710 4 US-08-957-310-6 Sequence 10, Appl
33 248 2.6 2366 1 US-08-480-604A-10 Sequence 10, Appl
34 248 2.6 2366 2 US-08-405-496A-10 Sequence 10, Appl
35 248 2.6 2366 4 US-08-915-136-10 Sequence 10, Appl
36 248 2.6 2366 4 US-08-957-310-10 Sequence 10, Appl
37 246.5 2.6 930 4 US-09-134-001C-5314 Sequence 5314, Ap
38 245 2.6 1566 2 US-08-687-996A-23 Sequence 23, Appli
39 244.5 2.6 2285 4 US-09-308-375-2 Sequence 2, Appli
40 243 2.6 1301 4 US-09-071-035-234 Sequence 234, App
41 243 2.6 1301 4 US-09-071-035-238 Sequence 238, App
42 243 2.6 1301 4 US-09-071-035-242 Sequence 242, App
43 242.5 2.6 1220 4 US-09-206-942-28 Sequence 28, Appli
44 242.5 2.6 1226 4 US-09-206-942-26 Sequence 26, Appli
45 241.5 2.6 1612 1 US-08-169-927-2 Sequence 2, Appli

```

ALIGNMENTS

RESULT 1

US-09-008-172-2

; Sequence 2, Application US/09008172

; Patent No. 6127602

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; TITLE OF INVENTION: Latexes in Paper Manufacture

; FILE REFERENCE: 0358D

; CURRENT APPLICATION NUMBER: US/09/008,172

; CURRENT FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/482,711

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-008-172-2

```

Query Match 26.0%; Score 2446; DB 3: Length 1430;
Best Local Similarity 46.7%; Pred. No. 1.1e+150;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

QY 738 LQINGQQYYIDPTGQPRKNFLQSGNNWYFSDTGVGTNALEQFAGTVSSNEQYR 797
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 174 VKQIDGKYYIIG-SDGPKKNFALTNNKVLVFDKNTGALTDTTSQYQFKGLTKLNDYT 232
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 798 NGAAYSYDDKSTENYNGYLTADTWYRPKQILKDGWTWDSKETDMEPILMYWNPQLQ 857
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 233 PHNOIVNFENTSLETIDNVYVTDASWYRPKDILKNGTWTASSESDLRPLMSWPKQYQ 292
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 858 AYLNMKQHG-----NLLPSALPFFNADAPAELNHYSEIVQONIEKRISGTNDWLR 912
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 293 IAYLNMNQQGLGTGEN-----YTADSSQESLNLAQTQVQVKIETKISQQTQWLR 344
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 913 TLMHDFVNNPMNMKQSENVNFSGIO--FGGFLKYNESDLTPYANSRLLGRMPIN-- 968
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 345 DIINSFVKTOPNWNSTESDTSAGEDHLQGGALLYNSDKTAYANSYRLLNRTTSQT 404
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 969 ----TKDQYRQGEFLANDIDNSPVQAEOLNLYLLNFGTITANNQANFDSVRV 1023
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 405 GKPKYFEDNSGGYDFLLANDIDNSPVQAEOLNLYLLNFGTITANNQANFDSVRV 464
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 1024 DAPNDIADLMTAQDYFNAAAGMD-SDAVSNKHINLELDNHNADPEYFNKIGNPOLTMD 1082
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 465 DAVDNVNADLQIASYLYKAHYGVKSEKNAIHNLSLEAWSNDPQYNKDKTGAQLPID 524
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 1083 DTIKNSLNHGLS-----DATN-----RNGLDAIVHQSLADRENNSTENVIPNYSFVRAHD 1133
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



QY 1601 FNGTNILHRGSGYVLKADGG-QYYNL-GTTTKQFLPIOLTGKKGKOGNFGVKGNDGN-YY 1657  
Db 1053 FNGTNILGRGVYVLKDNASDKYFELKGNQI--YLPKQMT--NKEASTGFV--NDGNGMT 1106  
QY 1658 FYDLAGNMYKNTFIEDSGVNGWYFEDQDQKMYENKHFVDVDSYGEKGYTFFLKNGVSFRGG 1717  
Db 1107 FYSTSGYQAKNSFYQDAKGNWYFEDNNGHMYGLQQLN---GE--VQYFSLNGVOLRES 1160  
QY 1718 LVQI-----DNGTYFEDNYGK-----MVRNOTINAGAMIYTLDEN 1752  
Db 1161 FLENADGSKNFGHLGNRYSNGYISFDNDKRWYFDASGVNAVGLKTINGTQYF--DQD 1218  
QY 1753 GKLIKASY 1760  
Db 1219 GYQVKGAW 1226

RESULT 3  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 26.0%; Score 2446; DB 4; Length 1430;  
Best Local Similarity 46.7%; Pred. No. 1.1e-150;  
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

QY 738 LOTINGQYYIDPTGQPKRNFLOSNGNWIYFSDTGTGVTNALLEQFAKGTVSSNEOYR 797  
Db 174 VQIDGKYYIG-SDGQPKNFALVNNKVLFDKNTGALDTSQYQKQGLTKLNNDYT 232  
QY 798 NGNAYSVDKSIENVNGYLTADTYRPKQILKDGTTWDSKETDMPILMYVWNTLTQ 857  
Db 233 PHNQIVNFENTSLETIDNYTADSWYRPRKDLKNGKTWTASSEDRLRLKASWPPDKQTQ 292  
QY 858 AYLYNMYKHG-----NLLPSALPENADADPAELNHYSEIVQQNIKRISETGNTDMLR 912  
Db 293 IAYLYNMQQGLGTEN-----YTDSSQESLNLAQTQVQVIERKISQTOQTQMLR 344  
QY 913 TLMHDFVTNPNMKNKDSNVNFSGIQ--FQGGFLKYENSDLTPPYANSYRLLRMPIN-- 968  
Db 345 DIINSVKTPQNNWSQTESDTSAGEKDHQGGALLYNSDKTAYANSYRLLNRTPTQT 404  
QY 969 -----IKDOTYRQGBFLANDIDNSPNVYQAEQLNWLYYLLNFGITANNQDANFDSVRY 1023  
Db 405 GKPKYFEDNSSGGYDFLLANDIDNSPNVYQAEQLNWLHYLMNYGSIVANDPEANFDGVRV 464

QY 1024 DAPDNIDADLMTIAQDYFNAAYGMD-SDAVSNKHINILEDWNHADPEYFNKIGNPOLTMD 1082  
Db 465 DAVDNVWADLLQASDYLLKARYGVDRKSEKNAINHLSLEAWSNDNDPOYNKDTKGAQLPID 524  
QY 1083 DTIKNSLNHGLS-----DAIN-----RWGLDAIVHQSLADRENKSTENVIPNYSFVRAHD 1133  
Db 525 NKRLSLYALIRPLEKRDASNKNEIRSGLEPVTNLSNRSARGKNSERANVYIFTRAHD 584  
QY 1134 NNSODQIQNAIR-DVTGK-DYHTFTFEDEQGDIDAYIQDONSTVKKYNYNIPASVAILL 1191  
Db 585 SEVQTVIAKIIKAOINPKTDLFTLDELQAKFIYNEQMRQAKKYTQSNITAFALML 644  
QY 1192 TNKDTIPRVYGDLYTDGGQYMEHQTRYDTLNLKSRVKYVAGGQSMQTSVGNNGN- 1250  
Db 645 SNKDSITRLYGDWYSDGQYMATKSPYDAIDTLKARIKAYAGGQDMKITTYVEDGKSH 704  
QY 1251 -----ILTSVRYGKGAMTATDGTDETGTQIGVVVSVNTPNLKLGVNDKVVHMGAAH 1303  
Db 705 MDWDYTGVLISVRGTGANEATDQSEATKTQGMAVITSNPSLKLQNDKVIIVNMGAAH 764  
QY 1304 KNOQYRAAVLTTTDDGVINYTSQDQAP--VAMTDENGDLYSSHNLVVNGKEEADTAVQGY 1361  
Db 765 KNOEYRPLLLTTKDLGTSYSDAAAKSLYRKTNKDELVEDASD-----IQGY 812  
QY 1362 ANPDVSGYLAWVPVGASDNQDARTAPSTEKNSGNSAYRTNAAAFDSNVIFEAFSNFYVTP 1421  
Db 813 LNPOVSGYLAWVPVGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNQDFV 872  
QY 1422 TKESERANVRIQNAADFFASLGFTSFEMAPQYNSKDRTEFLDSTIDNGYAFTRDYDGLMS 1481  
Db 873 TKSDYTNKKIAQNVQLFKSGVTSFEMAPQYVSSDGSFELDSIIQNGYAFEDRYDLAMS 932  
QY 1482 EPNKYGTDEDLRNAIOALHKAGLOVMADWPDQIYNLPGREVATVTRVDDRGVNWKDAIL 1541  
Db 933 KNNRYGQQDMINAVKALHKSQIOVIADWPDQIYNLPGREVATVTRVDDRGVNWKDAIL 992  
QY 1542 NNNLYVNTTIGG- EYQKYGAGFLDKLQKLYPIFTFKQVSTGVAIDSKQITWESAKY 1600  
Db 993 KNTLYAANTKSNKDYQAKYGGAGFLSELAARYPSIFNRTQISNGKIDPSEKITAMKAKY 1052  
QY 1601 FNGTNILHRGSGYVLKADGG-QYYNL-GTTTKQFLPIOLTGKKGKOGNFGVKGNDGN-YY 1657  
Db 1053 FNGTNILGRGVYVLKDNASDKYFELKGNQI--YLPKQMT--NKEASTGFV--NDGNGMT 1106  
QY 1658 FYDLAGNMYKNTFIEDSGVNGWYFEDQDQKMYENKHFVDVDSYGEKGYTFFLKNGVSFRGG 1717  
Db 1107 FYSTSGYQAKNSFYQDAKGNWYFEDNNGHMYGLQQLN---GE--VQYFSLNGVOLRES 1160  
QY 1718 LVQI-----DNGTYFEDNYGK-----MVRNOTINAGAMIYTLDEN 1752  
Db 1161 FLENADGSKNFGHLGNRYSNGYISFDNDKRWYFDASGVNAVGLKTINGTQYF--DQD 1218  
QY 1753 GKLIKASY 1760  
Db 1219 GYQVKGAW 1226

RESULT 4  
US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07

```

; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match          24.9%; Score 2347.5; DB 4; Length 1375;
Best Local Similarity 41.3%; Pred. No. 2.6e-144; Indels 175; Gaps 35;
Matches 538; Conservative 181; Mismatches 448;

Qy 549 SAVSGFDTTIKLTNDQOALNGQLVLLRFSKAADG--NPSGDNVTVDQFSKNYATTGNN 606
Db 29 TSLSGSLVKADSTDROQAVTTEQASLVTTSAAKETLTATDTSTATSATSQPTATVDN 88
Qy 607 FDYKVNQNYEFGSHATNOSDKDSQWIIIVLVNGKEVKRQLVNDTKEGAAGNRNDVY 666
Db 89 VS-----TNGS-----TNTANTA-----NFVV 107
Qy 667 KVPFAIENSSMGFGQITLTVTKNENVQLVHRFSNDVKTGEGNYV-----DFW 716
Db 108 KPTTTEQAKTNSDKIITTSKAVNR-----LTATCKFVPANNHTAHPKTVT 154
Qy 717 SELMPVKDSFGKNGPIKQ-----FGLOTNGQQYYIDPTTGQPRKNFL 761
Db 155 DKVIEPKPKI---GKLQPSLSQDDIAALGNVKNIRKVNKYYYY-KEDGTLOKRYAL 209
Qy 762 OSGNNWIFDSDTGVGTNALSQFAKGVSSNE---OVRNNGNAAYSVDKSIENNVGYLT 818
Db 210 NINGKTFEFDGALSNNLTSPK--KGNITNDNTNSFAQYQNYSTVDVANFEHVDHILT 267
Qy 819 ADTWPRQOILKDGTTWTSKETMRPILVWVWNTLTQAYLVNMQHGNLLPSALPFF 878
Db 268 AESWYRPRYILKDGKTWQSTKDFRPLMTWPDQETQOYVNMNAQLGIHOT---Y 323
Qy 879 NADADPAELNHYSEIVQONIEKRISGTGNTDLRLTLHDFVTNNPMKNKDSSENVFSGIO 938
Db 324 NTATSPQLNLAAQTQIKIEKITAENKTNWLRQTISAFVKTSQANSSEKPFDDHLQ 383
Qy 939 FQGGFLKYENSDLTPYANSDFYLLGRMPIN-----IKDQYRQGEFLLANDIDNSN 989
Db 384 -KGALLYSNSKLTSAQNSNRIILNPTNPTNOTGKKDPRYTADRTIGGYEFFLLANDVDNSN 442
Qy 990 PVQAEQLNLYLILNFGTITANNDQANFSDSVRYDAPDNIDADIMNIAQDYENAAVGM-D 1048
Db 443 PVQAEQLNWLHFLMNFNIYANDPDANFDSIRVDVNDVNDADLLQIAGDYLLKAAGIHK 502
Qy 1049 SDAYSKNHINTLEDWNHADPYFNKIGNPOLMTDDTKNSLNHGLSDATN-RWGLDAIVH 1107
Db 503 NDKAANDHLSLEAWSYNDTPYLHDDGDNMINMONRLRLSLLSLAKPLNORSQMNPLIT 562
Qy 1108 OSLADRENNSTENVIPNYSVRADHNNSDQIONAIR-----DVTGKDYHTTFEDEQK 1162
Db 563 NSLVNRDNDNATRAVPSYFIRAHSEVQDLIRNIIRLNPVNG---YSFTTEIEIK 619
Qy 1163 GIDAYIDQNSTVKYNNILNIPASYAILLTNKTDIRPVVYGLYTDGQYMEHOTRYDYD 1222
Db 620 AFEIYNKDLLATKYYTHYNALSYALLTNKSSVPVRYGDMFTDDGQYMAHKTINVEA 679
Qy 1223 LTNLLKSRVKKVAGQSQMOTMSVCGNNILTSVRYGKAMTATDTGDTFRTOGIGVVS 1282
Db 680 IETLLKARIKIVSGQAMRNQOV-GNSEIITSVRYGKALKATDTGRTTRTSGVAVIEG 738
Qy 1283 NTPNLKLGNDKVVLLHMGAAHKNQYRAAVLTITTDGVTINYTSDQAP--VAMTDENGDLY 1340
```

```

Db 739 NNPSLRLKASDRVVVNMGAHNAQAYRPLLLTDTNGIKAYHSDQRAAGLVRYTNDRGELI 798
Qy 1341 LSSHNLVWNGKEADTAVOGYANPDVSGVLAVVWPVGASDNQDARTAPSTENSGNSAYR 1400
Db 799 FTA-----AD--IKGYANPOVSGYLGWVWPVGAAADQDVRVAASTAPSTDGKSVH 846
Qy 1401 TNAAFDSNVIFEAFSNFVTPTKESERANVRIAQNAADFFASLGFTSFEMAPQYNSKDR 1460
Db 847 ONAALDSRVWFEGFSNFQAFATKKBEYTNVWIAKNVDKFAEWGVTDFEAPQYVVSSTG 906
Qy 1461 FLDSTIDNGYAFTRDVLGMSBPKNYGTDEDLRNAIQALHKAQLQVMAQVDPQIYNLP 1520
Db 907 FLDSVIQNGYAFTRDYLGLSKPNKYGTADDLVKAIKALHSGIKGYMAQVDPQYALPE 966
Qy 1521 KEVATYTRVDDRGNVKMDALINNNLVVN-TIGGGYQKQYGGAFLDKLOKLYPELFTKK 1579
Db 967 KEVYATRYVDKYGTPVAGSOIKNTLVVDGKSSGKQQAQYKAGFLEELQAKYPELPARK 1026
Qy 1580 QVSTGVAIDPSSQKITEWSAKYFNGTNILHRGSGYVLKADGGQYVNLGTTTKOFLPIQLTG 1639
Db 1027 QISTGVPMOPSVIKQWSAKYFNGTNILGRGAGYVLKQDQATNTYFSLVSDNTFLPKSLV- 1085
Qy 1640 EKKQNGEGFVKG--NDG-NYFYVDLAGNMVKNFTFIEDSVG-NWYFDDQDKVVENKHFVD 1695
Db 1086 NPHGTSSSVTGLVFGKGYIYSTSGNAQNAFI--SLGNWYTFDNNGYWYVYTGASIN 1143
Qy 1696 VDSYGEKGYIFFLKNGVSFRGLVQT-----DNGTY-----YFDNYGK 1733
Db 1144 -----GANYYFLSNGIQLRNAIYDNGKVSLSYNGDRRYENGYVLFQGWRYFON-GI 1196
Qy 1734 MVRNQIINAGAMIYTDENGKLIKASYNDAEYPTSTDVGKM 1775
Db 1197 MAVGLTRVHGAVOY-FDASG-----FOAKGQFITTAD-GKL 1230

RESULT 5
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6485203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```

Query Match          24.9%; Score 2347.5; DB 4; Length 1375;
Best Local Similarity 41.3%; Pred. No. 2.6e-144; Indels 175; Gaps 35;
Matches 538; Conservative 181; Mismatches 408;

Qy 549 SAVSGFDTTIKLTNDQOALNGQLVLLRFSKAADG--NPSGDNVTVDQFSKNYATTGNN 606
```









## RESULT 9

US-08-793-824-2  
: Sequence 2, Application US/08793824  
: Patent No. 5981838  
: GENERAL INFORMATION:  
: APPLICANT: Simpson, Christine Lynn  
: APPLICANT: Giffard, Philip Morrison  
: APPLICANT: Jacques, Nicholas Anthony  
: TITLE OF INVENTION: Genetic Manipulation of Plants to  
: INCREASE STARCH STORING CAPABILITY  
: NUMBER OF SEQUENCES: 2  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Griffith Hack & Co  
: STREET: Level 8, 168 Walker Street  
: CITY: No. 5981838th Sydney  
: STATE: New South Wales  
: COUNTRY: Australia  
: ZIP: 2060  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/793,824  
: FILING DATE:  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: AU PM7643  
: FILING DATE: 24-AUG-1994  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 61 2 9957 5944  
: TELEFAX: 61 2 957 6288  
: TELEX: 26547  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1577 amino acids  
: TYPE: amino acid  
: STRANDEDNESS:  
: TOPOLOGY: not relevant  
: MOLECULE TYPE: protein  
: ORIGINAL SOURCE:  
: ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 23.2%; Score 2190.5; DB 2; Length 1577;  
Best Local Similarity 38.0%; Pred. No. 5.3e-134;  
Matches 530; Conservative 210; Mismatches 474; Indels 181; Gaps 40;

Qy	438	KVTEKVNQLQ-----PQHQSIVSVRSFAD-----NGGNDKRDYDWF	477
Db	4	KVREKLKHKKNWVTIGVTLMSVALAGSLLAQGKVEADETSAPNGDGLQQLSE	59
Qy	478	PVILNOTASNIDTITMTSNGHLHAGWASONS-----NETPTAYAILNNGKEVTRQKM	531
Db	60	-----GRASLVTTTITVEQ-----ASASVSAVATASVSHETSFQAATSASVQEAQAQ	110
Qy	532	SLTARPDVAAYPSLYNSAVSGFTTTLKNDQVQALNGLOVLLRFSKAADGNPSGDNT	591
Db	111	TSPVASQEVAVSSQTQSSG-----QETQTEQVSQGTSTQVAGQTSQAQSTPS	158
Qy	592	VTDQFSKNYATPGDNEDYVKVNGQVEFSGWHATNQSDKDSQILVILVNGKEVKROLV	650
Db	159	VTEQARVRLTNAAPATATRAADSTIINANRNT-----ITITAGSTTPNVTTIT	209
Qy	651	--NDTKEGAGFNNDYKYNPAIENSSMSGFQGIITLPVTVKNNVQLVH--RFSNDVK	706
Db	210	GNTPPK-----PNVTVSPNGTRPNVTI-VTQPNQPNKPQPSQSPQPNK	253
Qy	707	TGEGNYVDVFWSELMPPVDSQKNGPLKQFGLQTINQQYVIDPTTQGPKNFLLSGNN	766
Db	254	PVQPNQPSL--DYKPVASN-----LKTIDGKQYVE--NGVVKKNAIELDR	297

Qy	767	WIVFSDTGVTGTTNALELOFAKGTVSSNEQYRNGNAAYSYDDKSIENVNGYLTADTWYRPK	826
Db	298	LYYED-ETGAMVDQSKPLYRAADAIENNSIYAVYNAQYDTSSKSFHLNFLTADSWYRPK	356
Qy	827	QILKGGTTWDSKETDMRPIILWMPNTLTQAYYLYNMQHGNLLPSALPFENADADPAE	886
Db	357	QILKDGKNWTASTEDYRPLMTWPPDKVTOVNYLYNASQQG---FGNKTYTTDDMSYD	412
Qy	887	LNHYSIVOQNTIEKRISSETGNTDLRLTLMHDFVTNNPNWNKDE--NVNFSGIFOGGFLK	945
Db	413	LAAAETVORGIEERIGREGNTITWLRLQMSDFIKTPQWNSSEDNLLVYKDHLOGGALT	472
Qy	946	YENSULTPYANDYRLGRLMPINIK-----DQYRGQEFLLANDIDNSNPVQAEOLN	998
Db	473	FLNNSATSHANSDFELMNRPTNOTGRKYHIDRSGGYELLANDIDNSNPVQAEOLN	532
Qy	999	WLYYLLNFGTITANNDQANFDSVRVDAPDNIDADLMNTAODYFNAAYG-DSDAVSNKHI	1057
Db	533	WLHYINNGISILGNDSANFSDGVRIDAVNDVADDLQTIASDFKERYVADNEAIAHL	592
Qy	1058	NLEDNHNADPEYFNKIGNPQTMDDTIKNS--LNHGLSDATRWGLDAIVHOSLADRENN	1116
Db	593	SILEAMSYNDHYNKDTKGAQLSDNPLRETLTTFLRKSNTYRGSLEVRITNSLNRSS	652
Qy	1117	STENVVPIPNYSFVRAHDNNSQDQIONAI--RDVTGK-DYHTFTFEDEOKGIDAYIODQNST	1174
Db	653	QKHTPRDANYIFVRADHSEVQAVLANISKQINPKIDGFTFTMDLQKAFEYINADIKA	712
Qy	1175	VKKYNYINIPASYAILLTNKTIPRYVYGDLYTDGQYMEHQTRYDYDTLNLLKSRKIV	1234
Db	713	DKKYTOYNIPIAAYATMLTNKDSITRVYGDLETFDDQYMAEKSPYNAIDALLRARIKV	772
Qy	1235	AGGOSMOTMSVGNNNILTSVRYGKAMTATDTGTDGTGIGVYVSTNPKLGVNDK	1294
Db	773	AGGDMKVTKLNG-YEIMSSVRYGKAEBANOLGTAEIRNQGNMLVTANRPMKLGANDR	831
Qy	1295	VYVHMGAAHKNQYRAAVLTDTGVTINTSDQAP---VAMTDENGDLVSLSHNLVYNGK	1351
Db	832	LVVNMGAHKNQAYRPLLSKSTGLATYLLKDSVPAGLVRYTDNOCNLTFTADD-----	885
Qy	1352	EEADTAVQGYANPDYSGYLAVWPVVGASNDQARTAPSTEKNSGNSAYTNAFQSNVIF	1411
Db	886	-----IAGHSTVEVSGYLAVWPVVGASENDQARTKASSTK-KGQVFESSAALDSQVYI	938
Qy	1412	EAFSN---FVYPTKESERANVRIAQNAOFFASLGFSEMAPOYNSSKDRFTLSDTIDN	1468
Db	939	EGFSNFQDFVKTPSQYTNRV---IAQNAKLFKEWGITSEFAPQVYSSODGTFLDSIEN	995
Qy	1469	GYAFTDRLGLHSEPNKYGTDEDLRNAIQALHAGLOVWADWVPDQIYNLPKQEVATYTR	1528
Db	996	GYAFEDRYDIAMSKNNKYGSLKDLMDALRALHABGISATADWVPDQIYNLPKQEVVTSR	1055
Qy	1529	VDRGNVWKDAIINNLYYVNT-IGGGEYQKYGGAFLDKLQKLYPEIFTKQVSTGVAI	1587
Db	1056	TNSYGTPRPNAEINSLYAAKTRTFGNDFOGKYGGGAFLDELKAKYPAIFERYOISNGRL	1115
Qy	1588	DPQSKTWSAKYFNQTNILHRSGGYVLKADG-QQYNNL--GTTTKQFLPOLTGEKKQG	1644
Db	1116	TTNEKITWSAKYFNQTNILHRSGGYVLKADG-QQYNNL--GTTTKQFLPOLTGEKKQG	1644
Qy	1645	NEGFKV-GNDGNYFYDLAGNNVKNFTIEDSVGNMFFDQDGKVENKHFVDVDSYGEKG	1703
Db	1170	GGGFRVGDVQY--LSIGGYLAKNTFIQVGNQWYFDFKNGNMTGEOVID----GKK-	1222
Qy	1704	TYFFLLKNYSFRGGL-----VQTDNGTY-----YFDNYGKWRNQ	1738
Db	1223	-YFFLDNGLQLRHVLRQSGDGHVYVDPKGVQAFNGFYDFAGPRODVRYFDNGQWYRGL	1281
Qy	1739	TINAGAMIYTLBENG	1753
Db	1282	HDMYGTFTFYDEKTG	1296

## RESULT 10

US-09-268-347-36  
; Sequence 36, Application US/09268347  
; Patent No. 6335182

## GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347  
; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 36  
; LENGTH: 2411  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-268-347-36

Query Match 3.5%; Score 327.5; DB 4; Length 2411;

Best Local Similarity 20.2%; Pred. No. 2.le-12;

Matches 418; Conservative 220; Mismatches 670; Indels 757; Gaps 106;

QY 11 KSGKOWTAAVATVAVSTALYGG---VAHADQQVQQAQTDDQSTVNN-----57  
Db 639 ENGKHTITVSAETRADSGLEKDGDTIKLVNDQ-----NTDNVLTGVNGTAVTKGF 692  
QY 58 DTKTVALD-----TNTDQSAQTDDKKQVVSNTNQSKTDDTSTADKKNSTSTPVSVLPSN 111  
Db 693 EVTKTGATDADRGKVTVDATANDAKK-----VATVKDVATA-INSAAFTV-----738  
QY 112 NTEQAKNYNEQDKNGYGNIDTAYFESNNQLHVGWGNATNASQGTNSRQIIVRDIITNNEL 171  
Db 739 KTENLTTSIDENPTDNGKDDALKAGDTLTFKAGNKKVVRDGN-----ITFDLAKNLEV 794  
QY 172 -----GRDVTNNVARP-----DVKNVHNY-----192  
Db 795 KTAKVSDTLTGNGPTGTTATPKVNTITADGLNFKAETADASGKNVYLKGIATLTLT 854  
QY 193 --NADNSGFDVNVNIDFKMKDYRDSIEIVSR-----YSGNGKSVDMWSQPIFEKNYAY 246  
Db 855 EFSAGAKSSHVDLNVDAIK-KSNAASIEDVLRAQWNIQGNVNDYVA---TYDVTNFTD 910  
QY 247 LDT-----FEVKNELHATGWNATNSAINYNHFFVI-----LFDOTNGKE 286  
Db 911 DSTGTTTIVTQKADGKADVKIG-----AKTSVVK-DHNGKLTGKDLKDANNAT 961  
QY 287 VARQEVREGQRPDVAKVYPQVGAAN-SGFNVTEINISDLDTYHOYQVLSRYSDNGE- 344  
Db 962 VSEDGKDTGTCLVTAKT---VIDAVNKSQWRTGEGATAE-TGATAV-----NAGNAET 1012  
QY 345 ---GDNVTWFPQSIAPANQSNQYLDSFDS-----KNGE-----VTVTGWNA 386  
Db 1013 VTSQTSVNFKNGNATTATVTSKDNNGINIKYDYNVGDGLKIGDKKIVADTTLTAVTGKV 1072  
QY 387 T---DLSELQNNHYTLFD--OTAKQVA-SAKADLSRPDVAKAYPTVKATNSGFKVT 440  
Db 1073 SVPAGANSVNNKKLVNAEGLATALNNISWTAKADKYAD---GESEGETDQEVKAGDKVT 1129  
QY 441 F-----KVNNTQPGHQYSVVSRF-----SADENGNGNKRHPDYWFSPIVLNQATSNIDT 490  
Db 1130 FRAGKNLKVQSEKDEKPTYSLODTLTGLTSITLGGTANGRNDTG-----TVINKDG 1179  
QY 491 ITMT-SNGLHAGWASDNSINETPYAILNNGKEVTRQKMSLTARPD---VAAYVPS 545  
Db 1180 LITLILANGA-AAGTASNGNTISVTKDGLSAGN-REITNVKSAIKTYKDTQNTAGATQPA 1237  
QY 546 LNSAVSGEDTTIKLTN-----DOYQALNGQLQVL---LRFKAADGNPS 587  
Db 1238 ANTAEVAKQD-LVLDTKPATGAAGADAKAPDTTAAATVGLRLGLWVLSAKKTDAD----1292  
QY 588 GDNVTVDQSKNYATITGGNFDIVKNGQVFSQGHATNQSDKDSQWILVLVNGKEVKR 647

Db 1293 -----ETQDKPEHAAVKNAVEFVFGKNGATVSAKTDN-----NGKHTVT 1332  
QY 648 QLVNDTKEGAAGFNNDVYKVPAINSSMSGFQGIITLPTVTVKNNVQLVHRFSNDVK 707  
Db 1333 IDVAEAKVG-DGLEKDTDGKIKLKVDNTD-----1360  
QY 708 GEGNYVDFWSELMPVKDSFOKMGPLKQFGLQTINGQQYVIDPTTGQPRKNFLLQSGNNW 767  
Db 1361 -----GNNLLT-----VDATKGAS-----VAKGEFNA 1382  
QY 768 IYFDSOTGVGTNALEQFAKQTVSSNEQVNGNAAYSDDKSTENVNGYLTADTWPRPKQ 827  
Db 1383 VTTDATTAQGTNAME---RGKVVVWKS--NGATATETDKKVVATVGDVAKA-----1428  
QY 828 ILKDGTTWDSKETDMRPILMVMWPNLTITQAYYLNMKQHGNLLPSALPFENADADPAEL 887  
Db 1429 -INDAATFVXVENDDSATI-----DDSP---1450  
QY 888 NHYSEIVQONIEKRISSETGNTDMLRLMHDFVT-----NNPMNKNKDSNFSGIQFGGF 943  
Db 1451 -----TDDGANDAKA--GDTLTKAGKNLKVVRDGNKNTIFA-----1485  
QY 944 LKENSOLTPYANSRYLLGMPINIKDQTYRGQEFLLANDIDNSNPVQVQAEOLNWLTYL 1003  
Db 1486 -----LANDLSVKSATVSDK-----1500  
QY 1004 LNFCTITANNDOANFDSVRVDAPDNIDADLNMIAQDYFNAAYGMSDASVSKHINILEDW 1063  
Db 1501 LSLGT---NGKNVNTS-----DTKGLNEAKD--SKTGDD---ANIHLNGIAS- 1540  
QY 1064 NHADPEYFNKIGNPQLTMDDTIKNS-----LNHGLSDATNR-----WGLDAI 1105  
Db 1541 -----TLDTLLNSGATTNLGGTITDNEKKRAASVKDYLNAGWNRGV 1584  
QY 1106 VHOSLADRENNSTENVVIPNVSVFRAHDNNSQDQIONAIRDVTG-KDYHTFTFE--DEQK 1162  
Db 1585 KPAS---ANNQVENI-----DFVATVD-----TVDFVSGDKDTTSVTVESKDNGK 1626  
QY 1163 GIDAYIQDQNSTVKKYLNIPASAYAILLNKDTIPRVYVYGDLYTGGQVMEHQTRYD 1222  
Db 1627 RTEVKIGAKTSVVKDHN-----GKLTGKE-----LKDANNNGVTVTE--TGKDE 1670  
QY 1223 LTNLKSRVYVA--GQSNOTMSVSGNNILTSVRYGKGAMTATDTG-----DE 1271  
Db 1671 NGULVTAKAVIDAVNKAGVRKVTGANGQNDPFAVAGSNVTFADGNGTAEVTRANDG 1730  
QY 1272 TRTOGIGVYVSNTPNLKLVNDKYLVEHGAHKNQYRAAVLTDTTQGVINYTSQOGAPVA 1331  
Db 1731 SITVKYNNKVAD--GLKLD-GDKIVAD-----TTVLTVADGKVT-----AP-- 1768  
QY 1332 MTDENGDLYLSSHNLVYVNGKEEADTAVQGYANPDVSGYLAWVPVVGASD--NODARTAPS 1389  
Db 1769 ---NNGD-----GKRFVDAS-----GLADALNKLSTATA 1795  
QY 1390 TERNSG-----NSAYRTNAAFDSNVIFEAFSNFYITPTKESERANVIAQNA-DFFASL- 1442  
Db 1796 GKEGTGEVDPANSAGQEVKAGD-KVTEKAGD-----NLKIKQSGKDFYSLK 1841  
QY 1443 ---GFTSFENAPQYNSKDKTFLDST--IDNGYAFTRDYLGLMSEPNKYCTDEDLNAI 1496  
Db 1842 KELDLTSVE---FKDANGGTGSEKTKITKDGTLITPANGAGAAGANTANTISVTKDG 1897  
QY 1497 QALHKA-----GLQVMADWVPDQIYNLPQKEVA-----TVTRVDDDRGNVWKDA 1539  
Db 1898 SAGNKAVTNVYVGLKFKFDG-----HTLANGTVADFEKHYDNAYKDLTNLDEKADNPT 1952  
QY 1540 IINNLYVNTIGG-----GEYQKYGAFDLKLOKLYPELFTTKQVSTGVAD 1988  
Db 1953 VADNTAATVGLRLGLGWVISADKTTGEPNOEYNAQVRNANE-----VKFKSNGINV- 2004  
QY 1589 PSQKITESSAKYFNGTNIL--HRGSGYVYK-----ADGGQ-----YYN-----1624  
Db 2005 -----SGKTLNGTRVITFELAKGEVKSNEFTVKNADGSEITNLKVKGVDMYISKEDID 2056

```
QY 1625 -----LGTTTKQFLP-----IQITGKKQ-----GNEGFKVGNNGYFYDLAGNMVKN 1668
Db 2057 PATSKPMTGKTEKYKGVKVSANGSKTEVTLNKGSGYVTGN-----QVADAIKAS 2109
QY 1669 TF---IEDSVGNWFFDQDKWE-----NKH-----FVDVDSY 1699
Db 2110 GFELGLADAAEAERAFSAESADKQLSKDKAETVNAHOKVRFANGLNTKVSATVESTIDAN 2169
QY 1700 GEKTYTFLKNGVSF-----RGLVQTDNGTYFYDNGYKVMYRNQTIINAGMIYT 1748
Db 2170 GDKVTTTFVKVDVPLPLTQIYNTDANGNKIVKADGKWELNADGTASKEVTILG-----N 2225
QY 1749 LDENG-KLIKASYN-SAEVPTSTD 1771
Db 2226 VDANGKKVKVTENGADKKWYITNAD 2250

RESULT 11
US-09-009-620-2
; Sequence 2, Application US/09009620A
; Patent No. 6127603
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357C
; CURRENT APPLICATION NUMBER: US/09/009.620A
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 349
; TYPE: PRN
; ORGANISM: Streptococcus mutans
US-09-009-620-2

Query Match 3.58; Score 326.5; DB 3; Length 349;
Best Local Similarity 26.5%; Pred. No. 1.4e-13;
Matches 102; Conservative 49; Mismatches 139; Indels 95; Gaps 12;

QY 549 SAVSGFDFTIKLTNDVQALNGQLVLRPFSKAADG--NPSGDMNTVTDQFSKNYATTGGN 606
Db 29 TSLSGSLVKADSTDDRQOAVTESQASLVTTSEAAKETILTATDTSTATSQPTATVTDN 88
QY 607 FDYKVGNGQVFEFGWHATNSDKDSQWIIIVLYNGKEVRQLVNDTKEGAAGFNNDVY 666
Db 89 VS-----TTNQS-----TNTTANTA-----NFVV 107
QY 667 KVNPAIENSSMGGQGIITLPTVTKNENVQLVHRFSNDVKTEGNYV-----DFW 716
Db 108 KPTTTSQAQKTDNSDKIITTSKAVNR-----LPTGKFPANNNTAHPKTVT 154
QY 717 SELMPVKDSFOGNGPLKQ-----FGLQTINGOQYYIDPTGQPKRNFLL 761
Db 155 DKVIEPKPI---GKLQPSLSQDDIAALGNVKIRKNGKYIY-KEDGTILQKNYAL 209
QY 762 QSCNNWIIFDSDTGVTGNALQLQAKGVSSNE---QVRNGNAAYSVDDKSIEHNWYLT 818
Db 210 NINGKTFEFTGALSNNTLFSK--KGNITNDNTNSFAQYQNYQVSTDVANFEHVDHYLT 267
QY 819 ADTWYRPOILKDGTTTWDTSKETDMRPLTMVWFWPNTLTQAVLYNMKQHGHLPLSALPFF 878
Db 268 AESWYRPYILKDKTWTQSTQKDEPRLMTWFWPDQETQRYQYVNMNAQLGIHOT-----Y 323
QY 879 NADADPAELNHYSEIVQONIEKRIS 903
Db 324 NTATSPLOLNAAQTQIKIEKIT 348

RESULT 12
```

```
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377.155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRN
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 3.4%; Score 321.5; DB 4; Length 2353;
Best Local Similarity 19.9%; Pred. No. 5e-12;
Matches 407; Conservative 222; Mismatches 640; Indels 775; Gaps 105;

QY 11 KSCQKWTTAAVATVAVSTALLYGG---VAHADQVQQAQSTTQQTSTVNN----- 57
Db 641 ENGKHITITVSVARTKADCGLEKDGDTIKLKVDNQ-----NTDNVLTGVNNGTAVTKGGF 694
QY 58 DTDKTVALD-----TNTDQSAQTDDKKQVSVNTNOSKTDSTDADKNSTSTPVSVLPSN 111
Db 695 ETVKTGATDADRGGKVTVKDATANDADKK-----VATVKDVATA-INSAAATFV----- 740
QY 112 NTEQAKNYNEODKNGYNIDTAYFNSNQLHYSGWGNATNASQGTNSRQIIVRDITNNEL 171
Db 741 KLENLTSIDEDNPTDNGKDDALKAGDTLTFRAGNMLKVKROGKN---ITFDLAKNLEV 796
QY 172 -----GRDVTNNVAP-----DVKNVHNVY----- 192
Db 797 KTAKVSDTLTIGTNTPTGTTATPKYNTITSTADGLNFAKETADAGSKNVYKLGIAITLT 856
QY 193 --NADNSGFDVNVNIDFSKMKYRDSIEIVSR---YSGNGKSVDMWWSQPIFDKKNVAY 246
Db 857 EPSAGAKSSHVDLVNDATK-KSNAASIEDVLRAGWNIQGNVNDIVA---TYDTVNFDT 912
QY 247 LDT-----FEVKNGLHATGWNATNSAINYNHHFVI-----LFDQTNGKE 286
Db 913 DSTGTTTIVTQKADGKGADVKIG-----AKTSVIK-DHNGKLTGKDLKDNANGAT 963
QY 287 VARQEVREGQSRPDVAKVYPOVVGAAAN-SGFNVTFNISDLDTYHQVLSRYSNSDNGE- 344
Db 964 VSEDDGKDTGTGLVTAKT---VIDAVNKGSRVGTGEGATAE-TGATAV-----NAGNAET 1014
QY 345 ---GDNVTVWFNPQSIAPANQSQGYLDSFDIS-----KNGE-----VTVTGWNA 386
Db 1015 VTSGSTSVNFKANGNATTATVSKDNGNINVKYDVNVGDLGKIGDDKKIVADTTLTITVGGKV 1074
QY 387 T---DLSELONNHYVILFD--QTAGKQVA-SAKADLIISRPDVAKAYPTVKTATNSGFKVT 440
Db 1075 SVFAGANSVNNKKLYNAEGLATALNNLSWTAKADYAD---GESEGETDQEVKAGDKYT 1131
QY 441 FKVNNLQPGHQYSVWSRSFADENGNDRHRHTDYFSPVILNQTASNIDITIM--TSNGL 498
Db 1132 FKA-----CKNLKV-----KQSEKDFTVS---LQDTLTGLTSITLGGTANGR 1170
QY 499 HIAGWMASDINSINETTFYAILNNGKEVTRQKMSLTARPDVAAVPSLVNSAVSGEDTTI 558
Db 1171 NDTG-----TVINKDGLTIT-----LANGAAAGTD--- 1195
QY 559 KLTDQYQALNGQLVLLRFESKAADGNPSGDNTVTVDQFS--KNYATTGGNFEDYVK----- 611
```

Db 1196 -----ASNGN-----TISVTKDGISAGNKEITNVKALKTYKDTQNTAETQDKP 1242  
QY 612 ---VNGQVFEFSGWHAATNOSNDKDSQWIIIVLNGKEVKRCLVNDTKEGAAGFNRNDVYKV 668  
Db 1243 AAVKNAEVEFVGKNGATVSATKDN-----NGKHTVTDVAEAKVG-DGLEKPDGKI 1294  
QY 669 NPALENSMSGFOGIIITLPTVTKNENVOLVHRFSNDVKTEGNYVDFWSELMPVKDSFOK 728  
Db 1295 KLVNDTD-----1302  
QY 729 GNGPLKQFGLQTINGOQYIIDPTGQPRKNFLQSGNNMIYFDSGTGVTNALELOFARK 788  
Db 1303 GNNLLT-----VDAIKGAS-----VAKGEFNAVTTDATTAQGTNAME-----RG 1341  
QY 789 TVSSNQYRNGNAAYSVDKSIENANGYLTDATWYRKPQILKDGTTWDSKETDMRPILM 848  
Db 1342 KYVYKGS--NGATATETDKKAVTGDVAKA-----INDAATFVKVENDDSATI-- 1388  
QY 849 VWPNTLTQAYLYNMYKQHGKLLPSALPPFNADADPAELNHYSEIYQONIEKRISGTNT 908  
Db 1389 -----DDSP-----TDDGAN 1398  
QY 909 DWLRTLMDHDEVT---NNPMWNKDSNVNPFSGIQFOGGFKLYENSDLTPYANSDYRLLR 964  
Db 1399 DALKA--GDTLTLKAGKNLVKXDKGNITFA-----1427  
QY 965 MPINIKDQTYRGOEFLLANDIDNSNPVVOAEQLNWLIIYLLNFGTITITANNDOANFDSVRVD 1024  
Db 1428 -----LANDLSVKSATVSDK-----LSLGT---NGKNVNITS-----1456  
QY 1025 APDNIDADLMNIAQDFYNAAGYDSDAVSNKHINILEDNHNADPEYFNKIGNPOLTMDOT 1084  
Db 1457 -----DTKGLNFARD---SKTGDD-----ANIHLNGIAS-----TLTDT 1487  
QY 1085 IKNS-----LNHGLSDATNR-----WGLDAIVHQSLADRENNSTENVIPNY 1126  
Db 1488 LLNSGATTNLGGGIIITDNEKKRAASVKDVLNAGWNYRKYKAS-----ANNQVENI-----1538  
QY 1127 SFVRAHNNNSQDQIONAIRVTC-KDYHTTFE--DEQKIDAYIQDQNSTVKKYKLYNI 1183  
Db 1539 DFVATYD-----TVDFVSGDKTTISVTESKONGRTEYKIGAKTSVKNDHN-----1585  
QY 1184 PASVAILLTNKDTIPRVYVYGDLYTDGQYMEHQTRYDTLTLLKSRVKYA---GGQSM 1240  
Db 1586 ---GKLTCKE-----LKDANNNGVTVTE--TDGKDENGVLTAKAYIDAVNKAQWRV 1633  
QY 1241 QTVSVGGNNLILSVRYGKGMATATDGT-----DETRTQIGVYVSTNPNLKLVN 1292  
Db 1634 KTTGANGQNDFAVATSGTNVTFADGNGTTAEVTKANDGSITVKNYKVAD--GLKLD-G 1690  
QY 1293 DKVVLHMGAAHKNQYRAAVLTITDGVINYTSQGPAPVAMTDGDLILSSHNLVYNGKE 1352  
Db 1691 DKIVAD-----TTVLTVADCKVT-----AP-----NNGD-----GKK 1717  
QY 1353 EADTAVGYANPDVSGYLAWVPVGSAD--NQDARTAPSTEKNSG-----NSAYRTNAAP 1405  
Db 1718 FVDAS-----GLADALNKLSTWATAGKEGTGEYDVPANSAGQEVKAG 1758  
QY 1406 DSNVIFAFSNFYVPTPKESERANVRITAOA-DFFASL-----GFTSFEMAPOYNSKDR 1459  
Db 1759 D-KVTFKAGD-----NLKIKSGKFTVLSLKKELKDLTSV-----FKDANGG 1800  
QY 1460 TFLDST--IDNGYAFTRYDPLGSEPNKYGTDEDLRNAIQALHA-----GLQVMADWY 1511  
Db 1801 TGSESTKITKDLTITPANGAGAAGANTANTISVTKGISAGNAKAVTNVSVGLKFKGDG- 1859  
QY 1512 PDQIYNLPKGEVA-----TVTRVDDRCNVKDKALINNLLYVNTIGG-----1553  
Db 1860 -----HTLANGTVADFEKHYDNYAKDITNLNDEKADNNPTVADTAATVGLRGLGWISA 1915  
QY 1554 ---GEYQKKYGGAFDLKQLKLYPEIFTKKQVSTGVAIDPSQKITEWSAKYFNGTIL-- 1607  
Db 1916 DKTTGEPNQEYNAQVRNANE-----VKFKSGNGINV-----SKTLNLTNRVITF 1959

RESULT 13

US-08-913-942-4  
; Sequence 4, Application US/08913942  
; Patent No. 6200578  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fiehr Hobbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,942  
FILING DATE: 29-DEC-1997  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,995  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/4031  
FILING DATE: 22-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Vance, Dolly A.  
REGISTRATION NUMBER: 39,054  
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2353 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-913-942-4

Query Match 3.4%; Score 321.5; DB 4; Length 2353;  
Best Local Similarity 19.9%; Pred. No. 5e-12;  
Matches 407; Conservative 222; Mismatches 640; Indels 775; Gaps 105;



QY 11 KSGKOWTAAVAVSTALLYG--VAHADQVQOASTTODQSTVNN----- 57  
Db 641 ENGKHTITVSAETKADCGLEKDGDTIKLVKDNQ-----NTDNVLTVGNNGTAVTRGGF 694  
QY 58 DDDKVALD-----TNTDQSAQTDDKKQVVSNTNQSKTDDTSTADKNKSTSTPVSPLSN 111  
Db 695 ETVKIGATDADRGKVTYRDATAADAKK-----VAIVKDVATA-INSAAIFV----- 740  
QY 112 NTEKAKNNEODKNGYKNIDIPAYSNNQLHVSNGWNAATNAGTNSRQIIIVRDIITNNEL 171  
Db 741 KTENLTTSIDEDNDPNDGKDDKALKAGDTLTFKAGKNLKVVRDGN-----ITFDLAKNLEV 796  
QY 172 -----GRDVTNNVARP-----DYKNVHNV----- 192  
Db 797 KTAKYSDILTICGNTPTGTTATPKVNITSTADGLNFAKETADASGKNVILKGIATLIT 856  
QY 193 --NANSQFVNINIDFSKMDYRDSIEIVSR-----YSGNGKSVSDWMSQITPDKNNYAY 246  
Db 857 EPSAGAKSHVDLNVDAITK- KSAASIEDVLRAGNIOGNGNVNHYVA---TYDTVNFTD 912  
QY 247 LDT-----FEVKNGLHATGNATNSAINYHFFVI-----LFDOTNGKE 286  
Db 913 DSTGTTTIVTOKAGKGADVKIG-----AKTSVIK-DHNGKLFTHGKDKDANNGAT 963  
QY 287 VARQEVREGQRPDAKVPYQVGAAN--SGFNVTENISDLDTYHQYQVLSRYNSDNGE- 344  
Db 964 VSEDDGKDTGTGLVTAKT---VIDAVKSGNRVTGEGATAE-TGATAV-----NAGNAET 1014  
QY 345 ---GDNVITWFNPQSIAPANQSNQYLDSPDIS-----KNGE-----VVTGUNA 386  
Db 1015 VTSQSVNFKNATATVSKDNGNINVDYVNGDGLKIGDKKIVADTTTLTVTGKV 1074  
QY 387 T---DLSELQNNHYVILFD--QTACKQVA--SAKADLISRPDVAKAYPVKATNSGPKVT 440  
Db 1075 SVPAGANSVNNKKLVNAEGLATALNLSWAKADKYAD--GESEGETDQEVKAGDVT 1131  
QY 441 FKVNILQPHQYQSVVSRSADENGNGDKRHTDYWFSVILNQASNIDITM--TSNGL 498  
Db 1132 FKA-----GKNUK-----KOSEKDTYS---LQDTLTGLTSLTIGGTANGR 1170  
QY 499 HIAGWASDNSINETTPYAILNNGKEVTRQKMSLTARPDVAAYVPSLYNSAVSGFDITI 558  
Db 1171 NDTG-----TVINKDGLTIT-----LANGAAGTD--- 1195  
QY 559 KLTNDQYQALNQLQVILRFSKAADGNFSGDNTVTDQRS--KNVATTGNGFDYK----- 611  
Db 1196 -----ASNGN-----TISYTKDGI SAGKEITNVKSAKTYKDTQNTADETQDKFPH 1242  
QY 612 ---VNGNQVFGSHATNOSNDKDWIIVLVNGKEVRQLVNDTKEGAAGFNNDVYKV 668  
Db 1243 AAVKANAEVEFGKNGATVSAKTDN-----NGKHTVIDYAEAKVG--DGLEKDTDKI 1294  
QY 669 NPAIENSSSGFGQIITILPVTYKNENVOLVHRFSNDVKTGEGNYVDFWSELNPKVDSQK 728  
Db 1295 KLVKDNTD----- 1302  
QY 729 GNGPLKQFGLQTIQGOYVYIDPTGPRKNELQSGNWIYFSDTGVTGVALELOPAK 788  
Db 1303 GNNLIT-----VDAIKAS-----VAKGEFNAVITDATTAGTNAE-----RG 1341  
QY 789 TVSSNEQYRNGNAAYSDDKSIENYNGYLTDATWYRQKILKDGTTWTDSDKETDWRPILM 848  
Db 1342 KYVWAKS--NGATATETDKKVVATVDVAKA-----INDAATFVKVENDDSATI-- 1398  
QY 849 VWNPNTLTCAYLYNMYKHQGNLLPALPFPFNADADPAELNHYSEIVQONIEKRSETGNT 908  
Db 1389 -----DOSP-----TDDGAN 1398  
QY 909 DWLRTLMHDFVT-----NNPMWNKDSENVNFSGIOFQGGFLKYENSDLTPYANSOYRLLR 964  
Db 1399 DAKKA--GDTLTKAGKNLKVVRDGNKITEA----- 1427  
QY 965 MPINIKDQTYRQGEPLFLANDIDNSNPVQAEQLNWLNYLLNFGITANNDQANFDSVRVD 1024

RESULT 14

US-09-669-974-33

; Sequence 33, Application US/09669974

; Patent No. 6353173

; GENERAL INFORMATION:

Db 1428 -----LNDLSVKSATVSDK-----LSLGT---NGKNVITS----- 1456  
QY 1025 APDNIDADLMAIADYENAAVGMDSDAVSNKHINILEDWNHADPEYFNKIGNQLMDDT 1084  
Db 1457 -----DTKGLNPAKD---SKTGDD---ANHLNGTAS-----TLTDT 1487  
QY 1085 IKNS-----LNGLSDATNR-----WGLDAIVHQSLADRENNSTENVIPNY 1126  
Db 1488 LLNSGATTLLGGNGITDNEKKRAASVXDVLNAGWNRVGPAS-----ANNQVENI----- 1538  
QY 1127 SFYRAHNNNSODOIQONAIRDVTG--KDYHTTFE--DEQKIDAYIODONSTVKYKLYNI 1183  
Db 1539 DFVATYD-----TVDFVSGOKDTTSVTVESKONGKRETKIGAKISVINDH----- 1585  
QY 1184 PASVAILLTKNDTIPRVYGYDLYTDGQYMEHQTRYDTLTLLKSRKYVA---GGQSM 1240  
Db 1586 ---GKLTGKE-----LKDANNNGVTVTE--TDGKDEGNGLVAKAVIDAVNKAGWRV 1633  
QY 1241 QTVSVGNNNLLTSVRYKKGAMTATDTGT-----DETRTQIGVYVVSNTNFKALGVN 1292  
Db 1634 KTTGANGQNDQFATVASGTVNITFADGNGTTAEVTKANDGSITVYKNYKVAD--GLKLD-G 1690  
QY 1293 DKVYLHGAHAKNQYRAAVLTITTDGVINTSDQGAPVAMTDENGDLYLSSHLNVVNGKE 1352  
Db 1691 DKIVAD-----TTVLTVADGKVT-----AP-----NNGD-----GKK 1717  
QY 1353 EADTAVOGYANPDVSGVLAVVVPVGASD--NQDARTAPSTEKNSG-----NSAYRANAAP 1405  
Db 1718 FVDAS-----GLADALNKLSTWATATAGKEGTGEYDVPANSAGQEVKAG 1758  
QY 1406 DSNVIFEAFSNFVPTKESERANVIAQNA-DFFASL-----GFTSFEMAPQYNSKDR 1459  
Db 1759 D-KVTFRAGD-----NLKIKSGKDTYSLLKELKDLTSE-----FKDANGG 1800  
QY 1460 TFLDST--INGVAFTRDYDLGMSHPKNTYDDEDLRNLIAOLHKA-----GLOWKADWV 1511  
Db 1801 TGSESTRITKDGTLTITPANGAGAAGANTANTISVTKDGISAGNAKAVTNVYVSLKFGDG- 1859  
QY 1512 PDQIYNLPGRKVA-----TVTRVDDRGVNVKDAIINNLYVYVNTIGG----- 1553  
Db 1860 ---HTLANGTVADFEKHYDNAYKDLTNLDEKGDNNPTVADNTAATVGDRLGLGWISA 1915  
QY 1554 ---GEVQKYGAFDLKLOKLYPEITFKQVSTGVAIDPSQKITEKSAKYFNGTIL-- 1607  
Db 1916 DKTTGEFQYNAQVRNANE-----VKPKSGNGIN-----SGKTLNGTRVITF 1959  
QY 1608 HRSGYVYK-----ADGGQ-----YYN-----LGTTKQFLP----- 1634  
Db 1960 ELAKGEVVKNEFTVKNADSGEINLVKVGDMYYSKEDIDPATSKPMGTGKTEKYKVGXV 2019  
QY 1635 IQLTGEKKQ-----GNEGFKVGDGNYFYFDLAGNVMKNTF-----IBDSVGNWYFFDQDGK 1686  
Db 2020 VSANGSKTEVTLTNKSGSYVTGN-----QVADATAKSGFELGLADAAEAFAESA 2072  
QY 1687 MVE-----NKH-----FVDVDSYGEKGTFFELKNGVSF----- 1714  
Db 2073 DKQLSKDKAETVNAHDKVRFANGLNFKVSAATVSESTDANGKVTTFVTVDLPLTQIY 2132  
QY 1715 -----RGLVQTDNGTYFYFDNYGKMWNRQNTINAGAMIYTLDENG-KLIKASYN-SDAEXP 1767  
Db 2133 NTDANGNKIVKADGKWYELNADGTASNKEVTLG---NYDANGKKVVKVTENGADKWY 2188  
QY 1768 TSTD 1771  
Db 2189 TNAD 2192

APPLICANT: PEAK, Ian Richard Anselm  
APPLICANT: JENNINGS, Michael Paul  
APPLICANT: MOXON, E. Richard  
TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
FILE REFERENCE: 065064/0128  
CURRENT APPLICATION NUMBER: US/09/669,974  
CURRENT FILING DATE: 2000-09-26  
PRIOR FILING DATE: 1999-08-19  
PRIOR APPLICATION NUMBER: PCT/AU98/01031  
PRIOR FILING DATE: 1998-12-14  
PRIOR APPLICATION NUMBER: GB 9726398.2  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 33  
LENGTH: 2353  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-669-974-33

Query Match 3.4%; Score 321.5; DB 4; Length 2353;  
Best Local Similarity 19.9%; Pred. No. 5e-12;  
Matches 407; Conservative 222; Mismatches 640; Indels 775; Gaps 105;

Qy 11 KSGKOWTAAVAVSTALLYGG--VAHADQQVQQAQTQDQSTTVNN-----57  
Db 641 ENGKTIIVSVAETKADGCKGDTIKLVNQ-----NTDNVLIVGNGTAVTRGGF 694  
Qy 58 DTKTVALD-----TMTDQSAQTTRKQVVSNTNQSDTDTSTADKNSTSTPVSPLSN 111  
Db 695 EVTKTGATDADRKGVTVDATANDAKK-----VATVKDVATA-INSATFV-----740  
Qy 112 NTEKAKYNEODKNGYNIIDTAIFSNQHLVSGHNATNASQINSROIIVRDIITNNEL 171  
Db 741 KTEINTTSLIDENPDNGKDDKALKAGDTLTFKAGKNLVKRDGN-----ITFDLAKNLEV 796  
Qy 172 -----GRDVTNNVARP-----DVKRNHNV-----192  
Db 797 KTAKYSDTLT IGGNTPTGTTATPKVNTSTADGNLFAKETADASGKNVYLKGIATLT 856  
Qy 193 --NADNSGFDVNVNIDFSKMDYRSIEIVSR---YSGNSKSDVMSQPIITFDKNVAY 246  
Db 857 EFSAGAKSHVDLNVDAK-KSNAASIEDVLKAGNIGQNGNVDYVA---TYDTVNFDT 912  
Qy 247 LDT-----FEVKNGLHATGMNATNSAINNHFYI-----LFDQTNCKE 286  
Db 913 DSTGTTTIVTQKADGKGADVKIG-----AKTSVIK-DHNGKLTGDKLKDANNCAT 963  
Qy 287 VARQEVREGQRPDVAKYIPQVVGAA--SGFNVTENISDLDTYHOYQVLSRYNSDNGE- 344  
Db 964 VSEDDGKDTGTGLVTAKT---VIDAVNKGSRVTEGATAE-TGATAV-----NAGNAET 1014  
Qy 345 --GDNVTYWFNPQSTAPANQSNQGLDSFDIS-----KNGE-----VTVTGWNA 386  
Db 1015 VTSQTSVNFKNATATVSKDNGNINVKYDVNVDGLKIGDDKKIVADTTLIVTGKV 1074  
Qy 387 T---DLSELQNNHYILED--QTAGQVA-SAKADLISRPDVAKAYPTVKVATNSGFKVT 440  
Db 1075 SVPGANSYNNKVLVNAEGLATALNLSWTAKADKYAD---GESEGETDQEVKAGDVT 1131  
Qy 441 EKVNNLQPGHOYSVYSRFSADENGNGDKRHTDYWFSVILLNQTSASNIDTITM--TSNGL 498  
Db 1132 FKA-----GKNLKV-----KQSEKDFIYS---LQDTLTGLTSLTGGTANGR 1170  
Qy 499 HIAGMASDNSINETTPYAILNNGKEVTRQKMSLTARPVAAYVPSLYNSAVSGFDFTI 558  
Db 1171 NDTG-----TVINKDGLTIT-----LANGAAAGTD---1195  
Qy 559 KLTNDQYQALNCQLVILRFSKADGNPSGNTVTDQFS--KNYATTGNGNDYK-----611  
Db 1196 -----ASNGN-----TISVTKDCISAGNKEITNVKSAUKTYKDTQNTADETQDKBEH 1242

Qy 612 ---VNGNQVEFSGWHATNQSDKDSQWIIIVLVNGKEVKRQLVNDTKREGAAGFNRRNDVYKV 668  
Db 1243 AAVKANANEVEFGKNGATVSAKTDN-----NGKHTVTIDVAEAKVG-DGLEKDDTGKI 1294  
Qy 669 NPAIENSSMSGFQGIITLPTVTVANENVQLVHFRSNDVKYKGEYVDFWSELMPVKDSFQK 728  
Db 1295 KLVQDNTD-----1302  
Qy 729 NGGPKQFGIQTINGOOYYIDPTTGOPRKNFLQSGNNMIYFSDRGVGTNALELOFAKG 788  
Db 1303 GNNLLT-----VDATKGAS---VAKGEFNAVTTDATTAQGTNAME---RG 1341  
Qy 789 TVSSNEQYRNGNAAYSDDKSIENGVNGILTADTWYRPKQILKDGTTWTDSKETDMPILM 848  
Db 1342 KVVYKGS--NGATATETDKKKVATVGDVAKA-----INDAAIFVKVENDDDSATI-- 1388  
Qy 849 VWPNTLTQAYLYLNMKQHGNNLLPSALFFPNADADPAELNHYSEIVQONIEKRISSETGNT 908  
Db 1389 -----DDSP-----TDDGAN 1398  
Qy 909 DWLRTLMHDFVT-----NPMWNKDSNVNPSGIGQFGGFLKYENSDLTPYANSDYRLGR 964  
Db 1399 DALKA--GDTLTLAGKNLKVKRDGNITFA-----1427  
Qy 965 MPINIKDQTYRGQEFLLANDIDNSNPVQAEQLNWLYYLLNFGTITANNDOANFDSVRVD 1024  
Db 1428 -----LANDLSVKSATVSDK-----LSLGT---NGKNVITS---1456  
Qy 1025 APDNIDADLNIADYFNAAYGMSDASVSNKHINILEDWNHADPEYFNKIGNPOLMDDT 1084  
Db 1457 ---DTPKGLNFARD--SKTGDD---ANIHLNGIAS-----TLTDT 1487  
Qy 1085 IKNS-----LNHGLSDATNR-----WGLDAIVHQSADRENNSTENVIPNY 1126  
Db 1488 LNSGATNLTGGNITDNEKKRAASVKDVLNAGNVGRVGPAS---ANNQVENI-----1338  
Qy 1127 SFVRAHDNNSQDQIQNAIRDVYG-KDYHTTFE--DBQKIDIDAYIQDNSTVKKYNLYNI 1183  
Db 1539 DFVATYD-----TVDFVSGDKDTSVTVESKONGKRTVKIGAKTSVIKDN---1585  
Qy 1184 PASVAILLTKNDTIPRVYIGDLYTDGGYMEHOTFYDYTLNLLKSRVKVA---GGQSM 1240  
Db 1586 ---GKLTQKE-----LKDANNNGVTVTE--TDGKDEGNGLYAKAVIDAVNKGARV 1633  
Qy 1241 QTSVSGNNNILTSVRYGKGAAMTATDTGT-----DETRTQIGVYVVSNTNPKLGVN 1292  
Db 1634 KTTGANGQNDDFATVASGTVNTEFADGNGTTAEVTKANDGSTTVKYNVKVD--GLKLD-G 1690  
Qy 1293 DKVYLHMGAAHKNOYRAAVLTITTDGVINYSDOGAPVAMTDENGDLYLSSHNLVYNGKE 1352  
Db 1691 DKIVAD-----TTVLTVADGKVT-----AP-----NGD-----GKK 1717  
Qy 1353 EADTAVQGYANPDVSGYLAVWVPYGASD--NQDARTAPSTEKNSG-----NSAYRTNAAF 1405  
Db 1718 FVDAS-----GLADALNKLSTWATACKEGTGEVDVPANSAGQEVKAG 1758  
Qy 1406 DSNVIFPAFSNVYTPKESERANVRITAOA-DEFFASL-----GFTSEFAPQYNSKDR 1459  
Db 1759 D-KVTFKAGD-----NLKIKQSGKDFTYSLKKELKDLTSVE---FKDANGG 1800  
Qy 1460 TFLDST--IDNGYAFTDRYDLGMSEPNKYGTDEDLRNAIOALHKA-----GLQVMADWV 1511  
Db 1801 TGSSTKITDKGLTITPANGAGAGAGATANTISVTKGISAGKNKAVTNVYSLGKKFGDG- 1859  
Qy 1512 PDQIYNLPKGEVA-----TVTRVDRGNWVKDAIINNLYYVNTIGG-----1553  
Db 1860 ---HTLANGTVADFEKHYDNAYKDLTLNDEKADNNPTVAADNTAAVGDRLGLGWISA 1915  
Qy 1554 ---GEYQKYGGAFLDKLQKLYPEIETFKQVSTGVAIDPSQKITEWSAKYFNGTIL-- 1607  
Db 1916 DKTTEEPNQEYNAOVRNANE-----VKFKSGNGINV-----SGTLNGTRVITF 1959  
Qy 1608 HRSGYVLK-----ADGGQ-----YYN-----LGTITTKQFLP-----1634

```
Db 1960 ELARGEVYKSNFTVKNRAGDGETNLVKVGDYYSKEDIDPATSKPMGTGKTKYKVENGVK 2019
Qy 1635 IQLRGEKKQ-----GNBGFVKGNDGNYFYDLAGNMVKNFT---IEDSVGNWFFDDGDK 1686
Db 2020 VSANGSKTEVTLTKNGSGYVTGN-----QVADAIKSGFELGLADAAEAERAFASAK 2072
Qy 1687 MVE-----NKH-----FVDVDSYBKGTYFFLKNGVSP----- 1714
Db 2073 DKQLSKDAETVNAHDKVRFANGNLNFKVSAATVESTDANGDKVTTTFVKTDVDELPLTQIY 2132
Qy 1715 -----RGLGVTDNGTYFYDNGKWMVRNQTINAGAMIYTTLDNG-KLIKASYN-SDAEYP 1767
Db 2133 NTDANGNIVKADGKWKVELNADGTASNKEVTLG---NVDANGKVKVKTENGADKWWY 2188
Qy 1768 TSTD 1771
Db 2189 TNAD 2192

RESULT 15
US-09-268-347-49
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-49

Query Match 3.4%; Score 318.5; DB 4; Length 2314;
Best Local Similarity 19.0%; Pred. No. 7.6e-12;
Matches 385; Conservative 248; Mismatches 736; Indels 655; Gaps 98;

Qy 1 MEIKKHFKLYKSGKOWTAA-----VATVAVSTALLYG---GVAHADQVQQ 44
Db 17 MAVAPCAKSHSGSSSTAGQVGSPPVIRLPRVATLAI---LVIGATLNGSAVAQNNSKI 73
Qy 45 ASTQDQSTVYNNDDTKVALDT-----NTDQSAQTDDKKQ 80
Db 74 AFGTGNNDNASASNEASIAIGSLAKAHANQAIAGGSKPPRNOAANQKAGSHAKGES 133
Qy 81 VYSNTNQSKTDDTSTA-----DKNSTP-----VSVLPNNTE-KQAKNYNEQDK- 125
Db 134 IAIGDVLAEGDASTAIGSDDLXLDNRNSTNSKYPNGLLSTLIQNHVLRQIRDSNGSOKY 193
Qy 126 ---GNVGNIDTA-----YFSNNQLHVSQWGNATNASQTSNRQIIVRDIITNNELGR 173
Db 194 RTAAEGHASTAVGAMAVAKHFAF-----AFGTRSTAEGNYSLAV---GLTAKAEKG 244
Qy 174 TDVTNNARPDPVKVHNY-YNADRSQGDVNVNIDFSKMKDYRDSIEIVSRYSNGKSVYD 232
Db 245 TIAIGS-----NAQAINYGALALGADTRVDLDYGIALGY--GSQILNNNNN----- 289
Qy 233 WSQPTTFPKNNYALDTEVKNGBELHATGNWATNSAINYHNFVILFDQTNKGEVAREV 292
Db 290 -----NNKAYVP-----EGNGSNIKSKATGNGLFSI-----GSSTIKRKII 326
Qy 293 REGQSRPDAKVPVQVGAANSGFNVFNISDLDTYHQYQVLSRYSNSDNGEGDNVYWF 352
Db 327 NVGAGYETDANNYAQLKAVE-----NLAKROI-----FKGDDNGTGK----- 366
Qy 353 NPQSTAPANQSNQGLDSFDSKNGEVTVTGWNATDISEL-QNNHYVILFDQTAGKQVAS 411
Db 367 -----KKLGSETLTIRGETQADKLTDNNDNNIGVYVDNNTGLVKL 405
```

```
Qy 412 AK----ADLISRPDV-AKAYPIVKTATNSGPKVTFKVNKLQPGH-QYSVVSFRPSADENG 465
Db 406 AKNLSLETVSTKNLTASEKVTVGSONNTA-----ELQSGGLTFTTTNASTDKTVY 457
Qy 466 GND-KRHTDYWFSPVILNOTASINIDTITSNGLHIAGWMASDNSINETTPYAILNNGK 524
Db 458 GTDGLAFTD-----NSNTALEDTTRITKDKI---GFSNKAGTVDENKPY----- 498
Qy 525 EVTRQKMSLTARPDVAAPVPSLYNSAVSGPDTITKLTNDQYQALNGLOLYLLRFSKAADG 584
Db 499 -LDKDKLKVG-----NSTLNNGLTV---NNTIGSNKQIOV-----GADG 535
-Qy 585 NPSEGDNVTVDQFSKNYATTGGNFDYVKVNGNOVEFSGHATNQSNKDKSQWILVNGKE 644
Db 536 IKFADYNNVNSNAKFGTT-----RITEEELGAD---ADGKVDKKSPLY---DKKQ 581
Qy 645 VKRQLVNDTKEGAAGFNRRND--VYKVNPAIENSSMSGFOGIITLPTVTVKNENVQLVHRES 702
Db 582 LQVGGVKITKD--SGINAGDQKISNYKDATDD-----TDAVYKQ----- 619
Qy 703 NDVKTBEGNVDWFSELMPVKDSFQKNGPLKQFGQTINGOQYYIDPTTGPQRKNFLQ 762
Db 620 -----LKQVQDADGALQSFSIRDEKGOEFTI-----SNLYS 651
Qy 763 SGNWIIYFSDTCVGTNALEL--QFAKGVSSNEQVRNGNAAYSDDKSIENVNGYLTPAD 820
Db 652 NGNTPNFTFITFAGENGISISNDIAKGVK-----VGIDPINGLTT--- 693
Qy 821 TWYRPKQIL---KDGTT-----WTDSKETDMRPILMVWVWPNLTLOAYLYNMYKQ 866
Db 694 ---PKLTVSGDKGTQLVIEQVASGNGTKNIIRGVSPTL-----PSITNAGGVRTTEQ 744
Qy 867 HGNLPSALPFFNADADPAELNHYSEIVQ--NIEKRISSETGNTDWLRLMHDFTVNNPM 924
Db 745 -GNTITS-----DEDKSAASIGDILATGPNLKNNSVGFVSTYNTV--DFIDGNAT 794
Qy 925 WNKDSENVNFSGIQFGGFLKYENSDLTPYANSDYELLGRMPINIKDQYRGEOEFLAND 984
Db 795 TAK-----VYDETQTSKVTD-----VNVEKIE-----LTGD 825
Qy 985 IDNSNPV-VOAEOLNWLYLLNEFTTANNDQANFDSVRVDAPDNID--ADLMNIAQDYF 1041
Db 826 NGATKIGVKITTLT-----TTNANGKATNFTSTDNALVNAKDAEENLTAKET 876
Qy 1042 NAAVGMDSDAVSNKHINILEDNWHADPEYF--NKIGNPQLTMDDTIKNSLNHGLSDATNR 1099
Db 877 HTTKGTADTALQI--FKVKKD-CATDEITVYCKDQTQNGKTVNTLKGKGLTVATNK 933
Qy 1100 -----WGLDAIVHOSLADRENNSTENVIPNYSFVRAHNNSSODQIQNAIRDVTGKDYET 1154
Db 934 DGTVTFGINTQSGLKAGDSTTLNKGDSIKNPA-----SNEQIQ-----VGADGVK 979
Qy 1155 FTFEDE---QKGIDA---YIQDQ-----NSTVKYNYLNIPASYALLTN 1193
Db 980 FAKVDKNSSTGIDGHSRITKDOIGFTGANGSLDTTKPHTLTKDKLVKGEVETNTGINAG 1039
Qy 1194 KDTIPRYVYGDLYTD-----GGQYMEHOTRYDYDTLNLKSRVKYVA--GGOSMOTMSV 1245
Db 1040 GKKITNIQSGDITQNSDAVTGG-----RVYDLKTE-LESKINSAKTAQNSLHEFSV 1091
Qy 1246 -----GNNNNILTSVRYGKGMATATDTGTDDETTOGI 1277
Db 1092 ADEQGNHFTVSNPYSDYTSKSDVTTFAGENGITTKVNGK-----VVRVGDIDTK--GL 1144
Qy 1278 GVVYSNTPNLKLGNDKVVLMHGAHKNQOYRAAVLTITTDG-----VINYTSDOG 1327
Db 1145 -----ITPKLTVG-----NNNGKGVIDSKDQGTITGLSNTLANVTND-G 1184
Qy 1328 APVAM-----TDEN-----GDLYLSHNLVYVNGEEDATVOGYANPDVSGYLAVWVP 1375
Db 1185 AGHALSQGLANDTKTRAASIGDVLNAGFNLOGNG--EAVDFVSTYDVTVDIDGNATTAK 1242
```

QY 1376 VGASD-----NODARTAPSTE-----KNSGNSAYRTNAAPDSNVIFE 1412  
Db 1243 VTYDDTSKTSVYDVNDVNDKTIETVSDKLGKVTITLTITSANGNATKPSAADGDALVK 1302  
QY 1413 AFSNFVYPTPKESBRANVRIAQNADFFASLGFTSFEMAPOYNSKDRFTFLDSTIDNCYAF 1472  
Db 1303 ASDIATHLNTLSGDIQTAKGASQASSAS-----YVDADGNKVIVDSTDKYYQV 1352  
QY 1473 TDRYDLGMSBNKYGTDEDLRNAICALHKAGLQYMAQVDPDIYNLPKKEVATVTRVDDR 1532  
Db 1353 NDK---GQVDRNKEVAKDL-----VAQAQTPD-----GTLAQMNVK 1386  
QY 1533 GNWVKDAIINNLLXVNTIGGGEYQKYG-----GAFLDKQLKLYPEIFTKKQVSTGVAID 1588  
Db 1387 SVIVKEQVNDAN-----KKQGINEDNAFIKGLENAAKDTKT---NAAVTVG 1430  
QY 1589 PSQITKWSAKYFNGTNIHLRSGYVLKADGQYNNLGTTKQFLPIQLTGEKKQGNEGF 1648  
Db 1431 DLNVAQPTLTFAGDTGTAKKLGETITIKGGQ-----TDNKL-----TDNNIGV 1476  
QY 1649 VKGNDGNYFYDLAGNMVKNFTIEDSVGNWYFFDQDGKMVENK--HFVDVDSYGEKGYF 1706  
Db 1477 VAGTDG--FTVKLAKDLT-----NLNSVNAGGTRIDEKGISFVDANGQAKANTPV 1524  
QY 1707 FLKNGVSFRG-----GLVQTDNGTYTFDNYGKAVR-----NQTINAGAMI--YTLDENG 1753  
Db 1525 LSANGLDLGGKRIISNIGAAVDDNDVNFQPFNEVAKTVNNLNQOSNGASLPFVVYTDANG 1584  
QY 1754 KLI-----KASYNSDAEY-----PTSTDVGKMLDONKL 1781  
Db 1585 KPINGTDGKPKOKAIGADGKYKHANANGVPYVDGDKPITDADKL 1628

Search completed: April 23, 2003, 15:11:19  
Job time : 55 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	9424	100.0	1781	9	US-09-995-749A-13		Sequence 2, Appli
2	2808	29.8	535	9	US-09-995-749A-2		Sequence 13, Appli
3	2446	26.0	1430	10	US-09-740-274-6		Sequence 6, Appli
4	2347.5	24.9	1375	10	US-09-740-274-4		Sequence 4, Appli
5	2276	24.2	1475	10	US-09-740-274-2		Sequence 2, Appli
6	1513.5	16.1	584	9	US-09-995-749A-12		Sequence 11, Appli
7	1512.5	16.0	522	9	US-09-995-749A-11		Sequence 11, Appli
8	1372	14.6	545	9	US-09-995-749A-10		Sequence 10, Appli
9	344.5	3.7	5795	10	US-09-815-242-12610		Sequence 12610, Appli
10	321.5	3.4	2353	10	US-09-797-862-33		Sequence 33, Appli
11	318.5	3.4	6281	10	US-09-815-242-12996		Sequence 12996, Appli
12	304.5	3.2	3158	10	US-09-815-242-12611		Sequence 12611, Appli
13	297.5	3.1	2434	10	US-09-815-242-5835		Sequence 5835, A
14	294.5	3.1	2086	10	US-09-815-242-5639		Sequence 5639, A
15	290	3.1	2344	10	US-09-815-242-12713		Sequence 12713, Appli
16	272.5	2.9	1349	10	US-09-815-242-5898		Sequence 5898, A
17	272.5	2.9	1349	10	US-09-815-242-13137		Sequence 13137, Appli
18	270	2.9	2122	9	US-09-813-214A-9		Sequence 9, Appli
19	266	2.8	1974	9	US-09-895-914A-12		Sequence 12, Appli

Db 181 ARPDKVNHVYNADNSGFDVNVNIDFSKMDYRDSIEIVSRYSGNGKSDVWNSQPTIFD 240  
QY 241 KNNYAYLDTFEVKGNGELHATGWNATNSAINTNHHFVILFDQTNKEVARQEVREGQSRPD 300  
Db 241 KNNYAYLDTFEVKGNGELHATGWNATNSAINTNHHFVILFDQTNKEVARQEVREGQSRPD 300  
QY 301 VAKVYPQVVGAAANSFNFTFNIISLDYTHQYQVLSRYSNSDNGEGDNVYWFNPSQTAPA 360  
Db 301 VAKVYPQVVGAAANSFNFTFNIISLDYTHQYQVLSRYSNSDNGEGDNVYWFNPSQTAPA 360  
QY 361 NOSNOGYLDSFEDISKNGEVTVTGWATDLSLQNNHYVILFDQTAGQVAKAKADLISR 420  
Db 361 NOSNOGYLDSFEDISKNGEVTVTGWATDLSLQNNHYVILFDQTAGQVAKAKADLISR 420  
QY 421 DVAKAYPVKTATNSGFKVTFKVNLPQCHQYSVYSRFSADENGNDKRRHTDYFSPVI 480  
Db 421 DVAKAYPVKTATNSGFKVTFKVNLPQCHQYSVYSRFSADENGNDKRRHTDYFSPVI 480  
QY 481 LNOTASNIDTTMTSNGLHIAGWASDNSINETTPYAILNNGKEVTRQKMSLTARPDVA 540  
Db 481 LNOTASNIDTTMTSNGLHIAGWASDNSINETTPYAILNNGKEVTRQKMSLTARPDVA 540  
QY 541 AVYPSLYNSAVSGPDTTILKLNDOYQALNGOLOVLLRFESKAADGNPSGDNVTVDQFSKNY 600  
Db 541 AVYPSLYNSAVSGPDTTILKLNDOYQALNGOLOVLLRFESKAADGNPSGDNVTVDQFSKNY 600  
QY 601 ATTGNGFDYKVNQGVFEFSQWHTNOSNDKDSQWIIVLVNGKEVKRQLVNDTREGAAGF 660  
Db 601 ATTGNGFDYKVNQGVFEFSQWHTNOSNDKDSQWIIVLVNGKEVKRQLVNDTREGAAGF 660  
QY 661 NRNDYKVNPAIENSMSGFGIITLPVTKNENVOLVHRESNDVKYGEKNYDFWSEL 720  
Db 661 NRNDYKVNPAIENSMSGFGIITLPVTKNENVOLVHRESNDVKYGEKNYDFWSEL 720  
QY 721 PVKDSFGKNGPLQFGLOTLNGQYYIDPTTGOPRKNFLQSGNNWYIFDSDTGCVGTNA 780  
Db 721 PVKDSFGKNGPLQFGLOTLNGQYYIDPTTGOPRKNFLQSGNNWYIFDSDTGCVGTNA 780  
QY 781 LELQFAKTVSNGQYRNGNAAAYSDDKSIENVNGYLTADTWYRKPQILKDGTTWTSKE 840  
Db 781 LELQFAKTVSNGQYRNGNAAAYSDDKSIENVNGYLTADTWYRKPQILKDGTTWTSKE 840  
QY 841 TDMRPILMWPNLTQAYLYNYWKQHGNNLPSALPFENADADPAELNHYSEIVQOINIEK 900  
Db 841 TDMRPILMWPNLTQAYLYNYWKQHGNNLPSALPFENADADPAELNHYSEIVQOINIEK 900  
QY 901 RISEGTNDWLRTLMHDFVTNNPMWKNKDSNVNFSGIFQGGFUKYENSDLTPYANSYDR 960  
Db 901 RISEGTNDWLRTLMHDFVTNNPMWKNKDSNVNFSGIFQGGFUKYENSDLTPYANSYDR 960  
QY 961 LLGRMPINIKDOTYRGOEFFLLANDIDNSNPVQAEQLNWLYYLLNFGTITANNDOANFDS 1020  
Db 961 LLGRMPINIKDOTYRGOEFFLLANDIDNSNPVQAEQLNWLYYLLNFGTITANNDOANFDS 1020  
QY 1021 VRVDAPONIDADLNNIAQDYFNAAAYGMDSDAVSNKHINILEDWNHADPEYFNKIGNPOLT 1080  
Db 1021 VRVDAPONIDADLNNIAQDYFNAAAYGMDSDAVSNKHINILEDWNHADPEYFNKIGNPOLT 1080  
QY 1081 MDDTIKNSLNHGLSDANRWGLDAIVHQSLADRENNSTENVVIPNYSFVRAHDNNSDQOI 1140  
Db 1081 MDDTIKNSLNHGLSDANRWGLDAIVHQSLADRENNSTENVVIPNYSFVRAHDNNSDQOI 1140  
QY 1141 QNAIRDVTGKDYHTFTFEDEQKIDAYIQDNSTVKYKNLYNPASVAILLTNKNDTIPRV 1200  
Db 1141 QNAIRDVTGKDYHTFTFEDEQKIDAYIQDNSTVKYKNLYNPASVAILLTNKNDTIPRV 1200  
QY 1201 YYGDLTYDGGQYMEHQTRYDTLTNLLKSRVKYVAGQSQMTSVSGNNNLLTSVRYGK 1260  
Db 1201 YYGDLTYDGGQYMEHQTRYDTLTNLLKSRVKYVAGQSQMTSVSGNNNLLTSVRYGK 1260  
QY 1261 AMTATDGTDETRTQIGVGVVSNTPNKLGVNDKVVLLHMGAAHKNQOYRAAVLTITDGI 1320

Db 1261 AMTATDGTDETRTQIGVGVVSNTPNKLGVNDKVVLLHMGAAHKNQOYRAAVLTITDGI 1320  
QY 1321 NYTSDGAPVAMTDENGDLYLSSHNLVWNGKEEADTAVOGYANPDVSGYLAVWVPYVAGSD 1380  
Db 1321 NYTSDGAPVAMTDENGDLYLSSHNLVWNGKEEADTAVOGYANPDVSGYLAVWVPYVAGSD 1380  
QY 1381 NODARTAPSTEKNSGNSAYETNAAFDSNITFEAFSNFVYTPTKESERANVRIAQNADFFA 1440  
Db 1381 NODARTAPSTEKNSGNSAYETNAAFDSNITFEAFSNFVYTPTKESERANVRIAQNADFFA 1440  
QY 1441 SLGFTSFEMAPOYNSSSKDRFTFLDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALH 1500  
Db 1441 SLGFTSFEMAPOYNSSSKDRFTFLDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALH 1500  
QY 1501 KAGLQWADWVWVPOIYNLPCKEYATVTRVDDRGWKNWDAIINNLLVYVNTIGGGEYQKXY 1560  
Db 1501 KAGLQWADWVWVPOIYNLPCKEYATVTRVDDRGWKNWDAIINNLLVYVNTIGGGEYQKXY 1560  
QY 1561 GGAFDLKLOKLYPEIFTKKQVSTGVAIDPSQKITESAKYFNGTNIHRGSGYVLKADGG 1620  
Db 1561 GGAFDLKLOKLYPEIFTKKQVSTGVAIDPSQKITESAKYFNGTNIHRGSGYVLKADGG 1620  
QY 1621 OYXNLGTTTKQFPIQIOTGEKKOGNEGFYKNGDGNYYFYDLAGNMVKNFTIEDSVGNWYF 1680  
Db 1621 OYXNLGTTTKQFPIQIOTGEKKOGNEGFYKNGDGNYYFYDLAGNMVKNFTIEDSVGNWYF 1680  
QY 1681 FDQDGRWVENKHFVDVDSYGEKGYTFELKNGVSFRGLVQTDNGTYTFDNYGKMYRNQTI 1740  
Db 1681 FDQDGRWVENKHFVDVDSYGEKGYTFELKNGVSFRGLVQTDNGTYTFDNYGKMYRNQTI 1740  
QY 1741 NAGAMIYTLDENGKLIKASYNSDAEYPTSTDVGMKMLDQNK 1781  
Db 1741 NAGAMIYTLDENGKLIKASYNSDAEYPTSTDVGMKMLDQNK 1781

RESULT 2

US-09-995-749a-13  
; Sequence 13, Application us/09995749a  
; Patent No. US20020155568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LOBBERT  
; APPLICANT: RAHACUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-995-749a-13

Query Match 29.8%; Score 2808; DB 9; Length 535;  
Best Local Similarity 100.0%; Pred. No. 2e-158;  
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 980 LLANDIDNSNPVQAEQLNWLYYLLNFGTITANNDOANFDSVRVDAPONIDADLNNIAQD 1039  
Db 1 LLANDIDNSNPVQAEQLNWLYYLLNFGTITANNDOANFDSVRVDAPONIDADLNNIAQD 60  
QY 1040 YFNAAYGMDSDAVSNKHINILEDWNHADPEYFNKIGNPOLTMDDTIKNSLNHGLSDATNR 1099  
Db 61 YFNAAYGMDSDAVSNKHINILEDWNHADPEYFNKIGNPOLTMDDTIKNSLNHGLSDATNR 120  
QY 1100 WGLDAIVHQSLADRENNSTENVVIPNYSFVRAHDNNSDQOIONAIRDVTGKDYHTFTED 1159

```
Db 121 WGLDAIVHQSADRENSTENVIPNYSFVRAHDNNSQDQIQNAIRDVTGKDYHTFTFED 180
Qy 1160 EOKGDAVIOQNSVVKYKYNLYNPASVAILLTKNDTIPRYVYGDLYTDGGCYMEHOTRY 1219
Db 181 EOKGDAVIOQNSVVKYKYNLYNPASVAILLTKNDTIPRYVYGDLYTDGGCYMEHOTRY 240
Qy 1220 YDTLNLKSRVYKYGQSMOTMSVGGNNILTSVRVYKGMATATDGTDETRTQIGV 1279
Db 241 YDTLNLKSRVYKYGQSMOTMSVGGNNILTSVRVYKGMATATDGTDETRTQIGV 300
Qy 1280 VVSNTPNKLGVDKAVLHMGAAHKNQOYRAAVLTDTTDDGVINYTSQOGAPVAMTDGDL 1339
Db 301 VVSNTPNKLGVDKAVLHMGAAHKNQOYRAAVLTDTTDDGVINYTSQOGAPVAMTDGDL 360
Qy 1340 YLSSHNLVYVNGKEEADTAVQGYANPDVSGYLAVVVPVPGASDNQDARTAPSTKNSGNSAY 1399
Db 361 YLSSHNLVYVNGKEEADTAVQGYANPDVSGYLAVVVPVPGASDNQDARTAPSTKNSGNSAY 420
Qy 1400 RYNAAFDSNVIFEAESNFVYPTKESERANVRIAQNAOFFASLGFTSEMAPOYNSSKDR 1459
Db 421 RYNAAFDSNVIFEAESNFVYPTKESERANVRIAQNAOFFASLGFTSEMAPOYNSSKDR 480
Qy 1460 TFLDSTIDNGYAFTDRYDLGMSKPNKYGTDEDLRNAIQALHKAQLOVMDVWPQ 1514
Db 481 TFLDSTIDNGYAFTDRYDLGMSKPNKYGTDEDLRNAIQALHKAQLOVMDVWPQ 535
```

## RESULT 3

```
US-09-740-274-6
; Sequence 5, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,520
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
```

## US-09-740-274-6

```
Query Match 26.0%; Score 2446; DB 10; Length 1430;
Best Local Similarity 46.7%; Pred. No. 1.9e-136;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;
```

```
Qy 738 LOTINGQOYIDPTTQPRKKNFLQSGNNWYFSDTGVGINALEQFAKGTVSSNEQYR 797
Db 174 VKQIDGKYYIG-SDGQPKKNFALTNNKVLFDKNTGALTDTSQYQKQGLTKLNNDYT 232
Qy 798 NGNAAYSDDKSIENGVYADTWYRPKQILKCDTWTWDSKETMRPILVWVWPNLTQ 857
Db 233 PHNQIVNENTSLIEDVNYVADSWYRPKQILKNGKTWASDESILRPILLSWVFDKQ 292
Qy 858 AYLLNMYKOHG-----NLLPSALPFNDADAPALNHYSEIVQOONIEKRISSETGNDWLR 912
```

```
Db 293 IAYLNYMNOOGLGTGEN-----YTADSSQESLNLAAQTVQVKLETKISQTOQTQWL 344
Qy 913 TLMHDFVTNPMNKNKSENVNFSIQ--FOGGLKYENSDLTPYANSDFYLGRLMPIN-- 968
Db 345 DIINSFKTOPNNSQTESDTSAGEKHLQGGALLYSNSDKTAYANSDFLLRTTTSQT 404
Qy 969 -----IKDQTYRGCEFLANDIDNSNPVQAEQLNMLYLLNFGTTTANDOANFVS 1023
Db 405 GKPKYFEDNSSGGVDFLLANDIDNSNPVQAEQLNMLYLLNMGYSIVANDPEANFDGVR 464
Qy 1024 DAPNDTADLMTAODYFNAAAGMD-SDAVSNKHINLEADNADPEYFNKICNPOLTMD 1082
Db 465 DAVDNVADLQASDYLKAHYGVDSEKNAIHLSTLEAWSNDNDPOYKDTKGAOLPID 524
Qy 1083 DTIKNSLNHGLS-----DATN-----RWGLDAIVHQSLADRENSTENVIPNYSF 1133
Db 525 NKLRSLLYALTPLKEDASKNKEIRSGLEPVTITSLNNSAEGKSERMANVIFTRAHD 584
Qy 1134 NNSODOIQONAIR-DVYTKK-DYHTFTFEDEQKGDAYIDQONSTVKKYNLYNPASVAILL 1191
Db 585 SEVOTVIAKIIRKQINPKTDGLTFTLDELKQAFKINVEDMRQAKKKTYSNITPATALML 644
Qy 1192 TNKDTIPRVYVYGDLYTDGGYMEHOTRYDPTLNLKSRVYKYGQSMOTMSVGGNN- 1250
Db 645 SNKDSITRLYVGYSDGQYMATKSPYDAIDTLKARIKAAAGQDMKITVYEGDKSH 704
Qy 1251 -----ILTSVRYGKGMATATDGTDETRTQIGVYVVSNTPNKLGVDKAVLHMGAAH 1303
Db 705 MDMDYTGLVTSVRYGTGANEATDQSEATKTQGNNAVITSNPNLKNQNDKVINMGAAH 764
Qy 1304 KNOQYRAAVLTDTTDDGVINYTSQOGAP--VAMTDENGDLYLSSHNLVYVNGKEEADTAVQGY 1361
Db 765 KNOEYRPLLTTKDGLTSYTSDAAAKSLYRKTNDKGLVFDASD-----IQGY 812
Qy 1362 ANPDVSGYLAVVVPVPGASDNQDARTAPSTKNSGNSAYRTNAAFDNSVIFEAESNFVYPT 1421
Db 813 LNPQVSGYLAVVVPVPGASDNQDVRVAAASNKANATQGYYESSSALDSQLIYEGFSNFODFY 872
Qy 1422 TKESERANVRIAQNAOFFASLGFTSEMAPOYNSSKDRTEFLDSTIDNGYAFTDRYDLGMS 1481
Db 873 TKSDYTNKKIAQNVOLFKNWGYTSFEMAPQYVSSDSELSLIQNGYAFEDRYDLAMS 932
Qy 1482 EPNKYGTDEDLRNAIQALHKAQLOVMDVDPQIYNLPGRKEVATVTVVDVDRGNVWKDAII 1541
Db 933 KNNKYGQSDMINAVKALHKSQIQVIADVWPQIYNLPGRKEVATVTVVDVDRGNVWKDAII 992
Qy 1542 NNNLYVYVNTTIGG- EYOKKYGGAFDLKLOKLYPEIFTKKOVSTGVAIDPSOKITEKSAKY 1600
Db 993 KNTLYAANTKSNKDYQAKYGGAFLSLAARYPSIFNRTQISNGKKIDPSEKITAMKAKY 1052
Qy 1601 FNGTNILHRSGVYVLKADGG-QYYNL-GTTTKOFLPLQLTGKKGQKQNEGFKVKGNDGN-YY 1657
Db 1053 FNGTNILGRGVYVLKDNASDKYFELKGNQT--YLPKQMT--NKEASTGFV--NDGNGMT 1106
Qy 1658 FYDLAGNMVKNTRFEDSVGNWYFFDQDGKMYENKHFVDVDSYGEKGYFFLKNKGVSPRG 1717
Db 1107 FYSTGYQAKNSFVQDAKGNWYFFDNNGHMYVYGLQQLN----GE--VOYFLSNGVQLRES 1160
Qy 1718 LVQPT-----DNGTYFFDNYGK-----MYRNQPTINAGMIYTLDEN 1752
Db 1161 FLENADGSKNYFGLHGNRYNGYYSFNDNSKRWYFDASGVMAVGLKTINGNTQYF--DQD 1218
Qy 1753 GKLIKASY 1760
Db 1219 GYQVKGAW 1226
```

## RESULT 4

```
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
```



Db	563	NSLVNRDNDNAETAAPVSPYSFIRAHDSHVQDLIRNIITEINPNVVG---YSFTTEBIKK	619
Qy	1163	GIDAYTQDQNSTVKKNYLNINIPASVAILLTNKDTPRYYVGLYTDGGOYMEHOTHRYD	1222
Db	620	AFEIYNKDLLATEKKYTHYNTALSYPALLTNKSSPRYYVGMETDGOYMAHKTINVEA	679
Qy	1223	LTNLLKSRVYVAGGOSMOTSGVGNINILTSVRVYKGGAMTATDITGDTFRTOIGVVVS	1282
Db	680	IETLLKARKYVSGGOAMENQOV--GNSEIITSVRVYKGGALKATDGTDRTRISGVAVIEG	738
Qy	1283	NTPLNLKLVGNDKVVLHMGAAHKNQOYRAVLTLTDDGVNTVNTSDGAP--VAMTDENGDLY	1340
Db	739	NNPSLRLLKASDRVYVNGAAHKNQAYRPLLLTDDNGIKAYHSDQEAAGLVRYTHNDRELI	798
Qy	1341	LSSHNLVYNGKEADPAVOGYANPDVSGYLAVVYVPGVSGDNODARTAPSTEKNSGNSAYR	1400
Db	799	FTA-----AD--IKGYANPQVSGYLVGVVWVPGVAAAQDQVRVAASTAPSTDGKSVH	846
Qy	1401	TNAAFDSNVIFEAFSFFVYTPTKESERANVRITAOANADFFASIGTSTSEMAPOYNSSKDRT	1460
Db	847	QNAALOSRVNFEFGSFFQAFATKKEYTNVYIAKNWDFAEWGYTDFEMAPQYVVSSTDGS	906
Qy	1461	FLDSTIDNGYAFDTRYDLGMSFNPKYGTDEDLRNAIQALHRAGLQVWADWPDQIYNLPG	1520
Db	907	FLDSVQNGYAFDTRYDLGISKPNKYGTADDLVKAIKALHKSIGKVWADWPDQMYALPE	966
Qy	1521	KEYATVTRVDDRGVWKKDAIINNLYVN--TIGGEYQKKYGGAFGLKLOKLYPEITTKK	1579
Db	967	KEYVATRVDDKGYTPVAGSQIKNTLYWDGKSSGKDQQAKEYGAFLEELQAKYPELFARK	1026
Qy	1580	QVSTGVAIDPSQKITWSAKYFNGTINLHRGSGYVLKADGGQYVNLGTTTKQFPLQIOLTG	1639
Db	1027	QISTGVPMPSVKIKOWSAKYFNGTINLGRGAGYVLKDAQNTVFSLSVSDNTFLPKSLV-	1085
Qy	1640	EKKQGNFVKV--NDG--NYPYDYIAGNMVKNFTFTDSVG--NWFFDQDGHMVENKHFVD	1695
Db	1086	NPNHGTSSSYTGLVDFGKGYVYTSYSGOAKNAFT--SLGNMYFYFDNNGYVMTGAQSIN	1143
Qy	1696	VDSYGEKGYFFFLKNGVSFRGLVOT-----DNGTY-----YEDNYGK	1733
Db	1144	-----GANYFYLSNGIQIRNAIYDNGKNVLSYQNDGRYVNGYVYLFQQQWRVFN-GI	1196
Qy	1734	MVRNQTINGAMITYLDELNGKLIKASYNDAEYPTSTDVGKM	1775
Db	1197	MAYGLTRVHGAVOY-FDASG-----FQAKGQFITAD-GKL	1230
RESULT 5			
US-09-740-274-2			
; Sequence 2, Application US/09740274			
; Patent No. US20020031826A1			
; GENERAL INFORMATION:			
; APPLICANT: Nichols, Scott E.			
; TITLE OF INVENTION: Glucan-containing Compositions and Paper			
; FILE REFERENCE: 0357CRD			
; CURRENT APPLICATION NUMBER: US/09/740,274			
; CURRENT FILING DATE: 2000-12-19			
; PRIOR APPLICATION NUMBER: 09/210,361			
; PRIOR FILING DATE: 1998-12-11			
; PRIOR APPLICATION NUMBER: 09/007,999			
; PRIOR FILING DATE: 1998-01-16			
; PRIOR APPLICATION NUMBER: 08/478,704			
; PRIOR FILING DATE: 1995-06-07			
; PRIOR APPLICATION NUMBER: 09/009,620			
; PRIOR FILING DATE: 1998-01-20			
; PRIOR APPLICATION NUMBER: 08/485,243			
; PRIOR FILING DATE: 1995-06-07			
; PRIOR APPLICATION NUMBER: 09/008,172			
; PRIOR FILING DATE: 1998-01-16			
; PRIOR APPLICATION NUMBER: 08/482,711			
; PRIOR FILING DATE: 1995-06-07			
; NUMBER OF SEQ ID NOS: 6			
; SOFTWARE: FastSeq for Windows Version 3.0			

```
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRN
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Qy 637 IIVNGKEVKQLVND-----TKEGAAGFNRNDVYKVNPAIENSSM 677
      : : : : : 24.2% Score 2276; DB 10; Length 1475;
      : : : : : Best Local Similarity 42.4%; Pred. No. 2.3e-126;
      : : : : : Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;
      : : : : :

Db 35 LVKADSNESKQISDNSTSVYTANEESNVITEATSKQEAASSQTN--HTVTSSTSV 92
      : : : : :
      : : : : :

Qy 678 SGFGIITLPVTKNENVQLVHRESNDVKTEGNYVDFWSEL-----MPVKD 724
      : : : : :
      : : : : :

Db 93 VNPKEVSNPTV--GETASNGEKLNQTTT-----VKTSEAAANNISKQTTADTDVID 146
      : : : : :
      : : : : :

Qy 725 SFQKNGEPL--KQFGLQINGQQYYIDFTQPKRKNFLOSNNWIVFD-----SDTGVG 777
      : : : : :
      : : : : :

Db 147 DSNAAANLQILEKLPNVEIDGKYYIYD--NNGKVRTNFTLIADGKILHFDGTGAYTDTSID 205
      : : : : :
      : : : : :

Qy 778 TNALELOFAKTVSSNQYRNGNAAYSDDKSIENVNGYLTADTWYRPKQILKDGTTWTD 837
      : : : : :
      : : : : :

Db 206 TVNKDIV-----TTSN-LYKYNQVDSQAQSFHVHYHTAESWYRPKVILKDKTWTQ 260
      : : : : :
      : : : : :

Qy 838 SKETDMRFLMWPNTLTQAYLYNMQHGNLPSLPFPNADADPAELNHYSEIYQON 897
      : : : : :
      : : : : :

Db 261 STERKDFRPLLTWTPDOETQRYVYNNAAQLGINKT---YDDTSNQLNIAAATQAK 316
      : : : : :
      : : : : :

Qy 898 IEKRISSETGNDLRTLMHDPVTNPMWKNDSNVNFSGTFQGGFLKYEN--SDLTPEYAN 956
      : : : : :
      : : : : :

Db 317 IEAKITLKNTDMLRQISAPVKQSAWNSDSEK--PFDD--HLQNGAVLYDNEGKLTPEYAN 374
      : : : : :
      : : : : :

Qy 957 SDYRLLRGMPIN-----IKDQTYRGQEFLLANDIDNSNPVVOAEQLNWLILNFG 1007
      : : : : :
      : : : : :

Db 375 SNYRILNRTPTNQKGGKDPRTADNTIGGYEFLANDVDNSNPVVOAEQLNWLHFLNFG 434
      : : : : :
      : : : : :

Qy 1008 TITANDQANSDSVRDAPONIDADLMNIAQDYENAAVGM--DSDAVSNKHINILEDNHA 1066
      : : : : :
      : : : : :

Db 435 NIYANDPDANDSDIRVADVDNVDADLLQIAGDYLKAAKHKKDAANDHLSLEAWSN 494
      : : : : :
      : : : : :

Qy 1067 DPEYFNKIGNPQMTMDTIKNSLNHGLSDATN--RWGLDAIVHOSLADRENSTENVVIPN 1125
      : : : : :
      : : : : :

Db 495 DTPYLHDDGDNNWINDKNLRLSLFLSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVS 554
      : : : : :
      : : : : :

Qy 1126 YSFVRAHDNNSQDQIQAIR-----DVTGKDYHTFTEDQKGDIDAYIODONSTVKKYNL 1180
      : : : : :
      : : : : :

Db 555 YSFIRAHDSEVQDLDIADIIKAEINPNVVG---YSFTMEIKKAEIYNKDLLATEKKYTH 611
      : : : : :
      : : : : :

Qy 1181 YNIPASVAILLTNRKDTIPRVYVGDLYTDGGQYMEHOTRYDITLNLKSRVKYVAGQSM 1240
      : : : : :
      : : : : :

Db 612 YNTALSYALLLTNKSVPVRYVYVDMFTDDGOYMAHKNTINVEAETLLKARIKYVSGQAM 671
      : : : : :
      : : : : :

Qy 1241 QTMVSGVGNNTLTSVRYKGGAMTATDGTDBTRTQGGVYVSNTPNPKLVNDKVLHMG 1300
      : : : : :
      : : : : :

Db 672 RNOQV--GNSEIITSVRYKGGALKATDGTDRTRTSGVAVIEGNNPSRLRKASDRVYVNMG 730
      : : : : :
      : : : : :

Qy 1301 AAHNQOYRAAVLTATDGVINVTSDQAG--VAMTDENGDIYLSHNLVVGKEADTAV 1358
      : : : : :
      : : : : :

Db 731 AAHNQAYRPLLLTDNGIKAYHSDQEAAGLVRYTNDRGELIFTA-----AD--I 778
      : : : : :
      : : : : :

Qy 1359 QGYANPDVSGYLAVWVPVPGAS--DNQDARTAPSTSEKSGNSAYRTNAAFNSNVIFEAFSN 1416
      : : : : :
      : : : : :

Db 779 KGYANPDVSGYLAVWVPVPGALIKWFAIRLA--RPHQQMASVHQAALDSRVMEFGFSN 835
      : : : : :
      : : : : :

Qy 1417 FVYPTKESERANVRIQAADFASLGFTSFEMAPQVNSKDRFTFLDSTIDNGYAFTRY 1476
      : : : : :
      : : : : :

Db 836 FOAFATKEEYTNVIAKNVDFEAWGVIDFEMAPQVVSSTGDSFLDSVQNGYAFTRY 895
      : : : : :
      : : : : :

Qy 1477 DLGSEPNKYGTDIEDNAIQALHAGLVQVADWPQIYNLPKGEVATVTRVDDRCNVW 1536
      : : : : :
      : : : : :

Db 896 DLGISKPNKYGTADLLYKAIKALSHKGIKVNADWPQMAFPEKEVVTATRVDDKYTPV 955
      : : : : :
      : : : : :

; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRN
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Qy 1537 KDAIINNYYVN--TIGGEYQKKYGGAFLDKLOKLYPELFTKKQVSTGVAIDPSOKITE 1595
      : : : : :
      : : : : :

Db 956 AGSQIKNTLYVVDGKSSGQQAQYKAGGAFLEELQAKYELFARKQISTGVPMDDPSVKIKQ 1015
      : : : : :
      : : : : :

Qy 1596 WSAKYFNGTNIHLRSGYVYLKADG--QYYNLTGTTTK--QFLPIQLTGEKKQSGNEGFVKGD 1653
      : : : : :
      : : : : :

Db 1016 WSAKYFNGTNIHLRSGAGYVYLKQATNTYFNISDNKEINFLPKTLLNODSQ--VGFSYDCK 1073
      : : : : :
      : : : : :

Qy 1654 GNYFYFDLAGNMVKNFTIEDSVGNWYFFQDCKMKWENKHFVDVDSYGEKGTTFLLKNGVS 1713
      : : : : :
      : : : : :

Db 1074 G-YVYVYSTGYQAKNFTISEG--DKWYFDNNGYMYMTGAQSIN-----GVNYFFLSNGLQ 1125
      : : : : :
      : : : : :

Qy 1714 FRGLVQJQITONGTY--YFDNYGKVMRN-----QTINAGAM-----IYTLDEN 1752
      : : : : :
      : : : : :

Db 1126 LRDAILKNEDGTAYVYNGDGRRYENGYQFMGVRHFHNGEMSVGLTVIDGQVQYFDEM 1185
      : : : : :
      : : : : :

Qy 1753 GKLIKASYNSDAEYPTSTDVGKM 1775
      : : : : :
      : : : : :

Db 1186 G-----YQAKGKFTTAD--GKI 1201
      : : : : :
      : : : : :

RESULT 6
US-09-995-749a-12
; Sequence 12, Application US/09995749a
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL--SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOUBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 584
; TYPE: PRN
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749a-12

Query Match 16.1%; Score 1513.5; DB 9; Length 584;
Best Local Similarity 53.2%; Pred. No. 8.3e-82;
Matches 312; Conservative 84; Mismatches 136; Indels 55; Gaps 14;

Qy 980 LLANDIDNSNPVVOAEQLNWLILNFGTITANDQANSDSVRDAPDNIDALMNAIQD 1039
      : : : : :
      : : : : :

Db 1 LLANDIDNSNPVVOAEQLNWLILNFGSITGNNDANFEDGIRVDADVNDVADLLKIAGD 60
      : : : : :
      : : : : :

Qy 1040 YFNAAYGMD--SDAVSNKHINILEDNHAHDPEYFNKIGNPQMTMDTIK----NSLNHGLS 1094
      : : : : :
      : : : : :

Db 61 YPKALYGTDKSDANANKHLSILEDWNGKDPQVYVNOGNAQLTMDYTVTSQFGNSLTHGAN 120
      : : : : :
      : : : : :

Qy 1095 DATNRW-----GLDAIVHOSLADREN-----NSTENVVIPKYSFVRAHDNNS 1136
      : : : : :
      : : : : :

Db 121 NRSNNMYFLDTGYLNGDLNKKIVDKNRPNSTGLVNRNANSGLTKVIPNYSFVRAHDYDA 180
      : : : : :
      : : : : :

Qy 1137 QDQIQNAIRD--VTGKDYHTFTEDQKGDIDAYIODON--STVKKYNLNINIPASVAILLT 1192
      : : : : :
      : : : : :

Db 181 QDPIKAMIDHGIKKWQDFTFTDQLAQGMEFYKQDENPSGFKKYNINDINLPSAYAMLTL 240
      : : : : :
      : : : : :

Qy 1193 NKDTIPRVYVGDLYTDGGQYMEHOTRYDITLNLKSRVKYVAGQSMOTMVSVG-----G 1247
      : : : : :
      : : : : :

Db 241 NKDTIPRVYVGDLYTDGGQYMEKGTIYNPVSALLKAR-KYVSGGOTMATDSSGDKDKG 300
      : : : : :
      : : : : :

Qy 1248 NNNILTSRYGKAMATATDGTDET----RTOGIGVYVSNTPNKLGVNDKVVYVLMHGAH 1303
      : : : : :
      : : : : :
```





Db	3029	RHQNATKQNL-----NTLSHLTNQKSAISSQDRACHGVSEVTAAKNAATELNTOMGN	3082
QY	1164	IDAYIQDQNSVKKYNLYNPASVAILLTNKTDIRVYVYGLYTDGQYME---HOTRYI	1220
Db	3083	LEQAIHQDN-TVKQ-----CVNFTDADKAKR---DAYTNVSAEETILNKTOGA	3127
QY	1221	DTLTLLKSRVYVAGGOSMOTMSVGGNNILTVSVRYKGGAMT-----ADTGT	1269
Db	3128	NFSKQDVEAIAQNVISAKN-----ALNGQONTNAKNTAKHALLNLTINNNAQKRLDTTKI	3183
QY	1270	DEFTRTGIGVYVSNT-----PNLKLGVNDKV-----VLHMGAAHKNOQYRAAVLTT	1315
Db	3184	DOATTIVAGVEAVSNTQTQINTAMANLONGINDKANTILASENYHDADSKKTAQTQAV-TN	3242
QY	1316	TGQVINYNTSDQCAPVAMTDE-----NGDLYL-----SSHNLVNVNGKEADTA--	1357
Db	3243	AENILNKSGSNLIDKAAVENALSQVYNAKAGLNGHNLEQAKSNANTIIINGHLITTAQK	3302
QY	1358	-----VQYANPD-----VSQYLAVVYVPGASDNODARTAPSTEKNSGNS	1397
Db	3303	DKLKQVOQAQNVAGVYDVTYKSSANTLINGAMGTLRNSIODNTATNNGQNYLDATESNKTN-	3361
QY	1398	AYRTNAEAFSNIYEEAFSNEVYPTKESERANVRIAQNADFFASLGFTSTEMAPOYNSKK	1457
Db	3362	--YNNAVDSANGVINATSN-----PNWDANAN-----QIATQVYSTK	3397
QY	1458	DRFLDSTIDNGYAFTRDYDLGMSBPNYKGTDEDLRNAIOA---LHKAGLQWADWYVPDQ	1514
Db	3398	NA---LDGT-----HNLTOAKOTATNAIDGATNLNKA---OKDALKAQ	3434
QY	1515	IYNLPCKEYATVTRVDDRGNVWKDAIINNLYVNTIGGGVQYKYGAFDLKLQKLYPE	1574
Db	3435	VTS---AQRVANVTSIOOTANELNTA-MGLOHNGIDDENATQTQKYRDAEQSKTAYDOA	3491
QY	1575	IFTKQV-----STGVAIDPS-OKITEWSAKYFNGTNIILHRSGYVYLKADGGQYNNL	1625
Db	3492	VAAAKAILNKQTSNSDKAAVDRALQOVT-----STKDALNGDAKLAEAKAAAKQNL	3543
QY	1626	GTITTKQFLPIQLTGKCKOGNEGFVKGN--DGNVYFYDLAGNMVK--NTF-----IED	1673
Db	3544	GTILN-----HITNAORTALEQINQWATVDG-----VNTVKTNTANLIDGAMNSLQK	3589
QY	1674	SVGNWTFYDQDGKKWENKHFVDVDSYGEKGYVFFLKNGVSPRGLGYVDNG	1724
Db	3590	SIN-----DKDAPL-RNQNYLDADE-SKRNAV---TQVATAEGLNKQTG	3630

## RESULT 10

US-09-797-862-33

; Sequence 33, Application US/09797862

; Patent No. US20020102276A1

; GENERAL INFORMATION:

APPLICANT: PEAK, IAN RICHARD ANSELM

APPLICANT: JENNINGS, MICHAEL PAUL

APPLICANT: MOXON, E. RICHARD

ALLEGRI, ROBERTO E. RICHARD  
: TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

FILE REFERENCE: 065064/0134

FILE REFERENCE: 005004/0134  
: CURRENT APPLICATION NUMBER: US/09/797.862

: CURRENT FILING DATE: 2001-05-03

; CURRENT FILING DATE: 2001 03 03  
 : PRIOR APPLICATION NUMBER: PCT/AU00/00000

: PRIOR FILING DATE: 1998-12-14

PRIOR FILING DATE: 1998-12-14  
PRIOR APPLICATION NUMBER: GR 9726398 2

; PRIOR APPLICATION NUMBER: GB 9/2000  
 : PRIOR FILING DATE: 1997-12-12

; PRIOR FILING DATE: 1997-12-12  
: NUMBER OF SEQ ID NOS: 33

```

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Paton1a v0.3.1

```

; SOFTWARE: PAT  
; GEO ID NO 33

```

; SEQ ID NO 33
; LENGTH: 2353

```

; LENGTH: 2353  
; TYPE: CFETYPE: PRT  
ORGANISM: H3

### Query Match

3.48; Score 321.5; DB 10; Length 2353;

Best Local Similarity 19.9%; Pred. NO. 1.7e-10;  
Matches 407; Conservative 222; Mismatches 640; Indels 775; Gaps 105;

Qy	11	KSKQWTVRAVAVSTALLYGG--VAHADQOQVQARSTQDOTSTVNN-----	57
Db	641	ENKHHITVSVAETKACDGLKGDTIKLKVDNO-----NTDNVLTVGNGNTAVTKGCF	694
Qy	58	DTDKTVALD-----TNTDQSAQTTDKKQVVSNTNQSKTIDTSTADKNSTSTPVSJLPSN	111
Db	695	ETVKTGATDADRGKVTVKDATANDAKK-----VATVKQVATA-INSAAATFV----	740
Qy	112	NTEKQAKNYEODKGYNIDTAYFSNNOLHVSGWGNATNASOCTNSROIIVRDIITNNEL	171
Db	741	KTENLTISIDEDPTDNGKDDALKAGDTLTFKAGKULKVKRDKN-----TFDIAKLKLEV	796
Qy	172	-----GRDVTNNVARP-----DVKNVHNY-----	192
Db	797	KTAKVSDTLTIGGNTPTGGTTATPKVNIITSTADGLNFAKETADAGSKNVYUCLKIATLTL	856
Qy	193	--NADNSGFDVNNIDFSKMKQYRDSIEIVSR-----YSGNGKSVDMWNSOPTFOKNYAY	246
Db	857	EPSAGAKSSHVDLNVDAIK-KSNAASIEDLVLAGNIOGNGNVDYVA---TYDTVNFDT	912
Qy	247	LDT-----FEVNGELHATGWNATNSALNYHHFVI-----LFDQTNGKE	286
Db	913	DSTGTTTVTVIQADKGADVIG-----AKTSVTK-DHNGKLF7GDKLDKANNGAT	963
Qy	287	VAROEVRGOSRPDAKVYPOVYGAAN-SGFNVVTENISDLDTYHOYQVLSRYSNDSNGE-	344
Db	964	VSEDDGKDTGTGLVTAKT---VIDAVNKSGRVTCGATAE-TGATAV-----NAGNAET	1014
Qy	345	--CDNVTYWFNPQSTAPANOSQGYLDSFDLS-----KNGE-----VNVTCNA	386
Db	1015	VTSGTSVNFKNGNATTATVSKONGINWYDVNGDGLKIGDOKKIVADTTLTVTGKV	1074
Qy	387	T---DUSELQNNHYILFD--QTAKQOVA-SAKADLISRPDYAKAYPVVKATNSGPKVT	440
Db	1075	SVPAGANSVNNKKLYAEGLATALNLSWTAKADKYAD--GESEGETDOEVGAKGKVT	1131
Qy	441	FKVNNLOPGHOYSVWSRESADENGNGDKRHHDYWFSPVILLNOTASNIDTITM--TNGNL	498
Db	1132	FKA-----GRNLKV-----KQSKDXFYIS---LODTLTLGLTSITLGTANGR	1170
Qy	499	HIAGWASDMSINETTPYAILNNGKEVTRQKMSLTARPDVAAYVPSLYNSAVSGFDTTI	558
Db	1171	NDTG-----TVINKDGLTIT-----LANGAAAGTD-----	1195
Qy	559	KLINDQYALNGQOLLRLRFSAADGNPSGDNVTVDQFS--KNVATTGCGNDYVK-----	611
Db	1196	-----ASNGN-----TISYTKDGISAGNKEITNVKSALKTYKDTQNTADETDKEEH	1242
Qy	612	---VNGNOVEFSGWHATNOSDKDSOWIIVLVNGKEVRQLVNDTKEGAAGFNNDYVK	668
Db	1243	AAVKNAREVEFVGKNGATVSAKTDN-----NGKHTVIDVAEAKVG-DGLEKDDTGKI	1294
Qy	669	NPAIENSMSGFGOITLPTVTKNENVQLVHRFSNDVKTGEGVYVDFVSELMVPKDSFQK	728
Db	1295	KLKVDNTD-----	1302
Qy	729	GNGLPKFGGLQTINGQOQYIDPTTGOPKXNFLLQSGNNWIYFSDTGYCTNALELOFAGK	788
Db	1303	GNNLLT-----VDA7FKGAS-----VAKGEFNAVTTDATAGTQWANE-----RG	1341
Qy	789	TVSSNEOYRNCNAAYSYDDKSIENVNGYLTDWYRPKQILKDGTTWDSKRETDMPFILM	848
Db	1342	KVVYKGS--NGATATETDKKATVGDVAKA-----INDAATFVKVENDSDSATI--	1388
Qy	849	VWVENTLITQAYILNYMKQHNLLPSALPFENADADPAELNHVSEIVQONIEKR1SEVGN	908
Db	1389	-----DDSP-----TDDGAN	1398
Qy	909	DWLRTLMHDEVT-----NNPMWAKDSENVNFSIQGQGF7KYENSDLTPYANSYRLLGR	964

Db 1399 DALKA--GDTLTLKAGNNLKVKRGKNTIFA----- 1427  
Qy 965 MPINIKDQYRQGEFLLANDIDNSNPVYQAEQLNWLNYLLNFGTITANNDOANFDSVRVD 1024  
Db 1428 -----LANDLSVKSATVSDK-----LSLGT---NGNKVNITS----- 1456  
Qy 1025 APDNIDADLMIAODYNAAVGMDSDAVSNKHINILEDWNRHADPEYFNKIGNPOLIMDDT 1084  
Db 1457 -----DTKGLNFADK-----SKTGDD-----ANIHLNGIAS-----TITDT 1487  
Qy 1085 IKNS-----LNHGLSDATR-----WGLDAIVHQSLADRENSTENVVIPPY 1126  
Db 1488 LLNSGATINLGGITDNEKKRAASVDVNLNAGWNRGVKPA-----ANNOVENI----- 1538  
Qy 1127 SFVRAHDNNSODQONAIIRDVTG-KDHTFTFE--DSQKGDIAVIOQNOSTVKKYNYNI 1183  
Db 1539 DFVATYD-----TVDFVSGDKDTSVTVESKDKRGKTEVKIGAKTSVIKDN----- 1585  
Qy 1184 PASVAILLTNKDTIPRVYIGDLYTDGGQYMEHQTRYDITLNLKSRVKYVA-----GGOSM 1240  
Db 1586 -----GKLFTEK-----LKDANNNGYTVIE--TDGKDEGNGLVTAKAVDAVNKAGWRV 1633  
Qy 1241 QTMVGGNNNLTLSVRGKGAMTATDGT-----DETRTOGIVVVVNTPNKLGLVN 1292  
Db 1634 KTTGANGQNDFAVAGSTNVTFADGNGTTAEVTKANDGSIYKYNKVKAD--GLKLD-G 1690  
Qy 1293 DKVYLHMAAHKNOYRAAVLTITDGVINYSOGAPVAMTDENGDLXLSHNLVYNGKE 1352  
Db 1691 DKIVAD-----TTVTVVADGKVT-----AP-----NNGD-----GKK 1717  
Qy 1353 EADTAVOQYANPDVSGYLAWVPVYGASD--NQDARTAPSTEKNSG-----NSAYRTNAAF 1405  
Db 1718 FVDAS-----GLADALNKLSTWATAGTEGTEVDPANSAQGEVKAG 1758  
Qy 1406 DSNVIFEAFNSVYTPKESERANVRIAQA-DFFASL-----CFTSEKAPQYNSSKDR 1459  
Db 1759 D-KVTFRAGD-----NLKIQSGKDFYSLKELKDLTSVE-----FKDANGG 1800  
Qy 1460 TFLDST--IDNGYAFTRIDYDLGMSNPKNYGTDEDLRNAIQALHKA-----GLQWADWV 1511  
Db 1801 TGSESTKITKDGITTPANGAGAAGANTANTISVTKDGISAGNKAATNVVYSGLKKEGDD- 1859  
Qy 1512 PDQYNLPCKEVA-----ADGQO-----YYN-----LGTITKQPLP----- 1634  
Db 1860 -----HTLANGTVADFEKHYDWAYKDLTNLDEKADNNTVADNTAAVGDRLGLGWISA 1915  
Qy 1554 -----GEYQKYGGAFLDKLQKLYPEITFKQVSTGVAIDPSQKITESAKYFNGTNIL-- 1607  
Db 1916 DKTTGEPNOEYNAQVRNANE-----VKPKSGNGINV-----SGKTLNGTRVITP 1959  
Qy 1608 HRGSGYVLK-----ADGQO-----YYN-----LGTITKQPLP----- 1634  
Db 1960 ELAKGEVYKSNFTVKNADGSETNLVKYGDMYKSKEDIDPATSKPMTGKTEKYKVENGV 2019  
Qy 1635 IQLTGERKKQ-----GNEGFAKNGDNGYFYFLAGNMVKNTF--IEDSVGNWFFDODGK 1686  
Db 2020 VSANGSKTEVTLTNKSGSYVTGN-----QVADAIKSGFELGLADAAFAKFAFASAK 2072  
Qy 1687 MVE-----NKH-----FVDVDSYGEKGYFFLKNGVSF----- 1714  
Db 2073 DKQLSKDKAETVNAHDKVRFANGINTKVSAAVTESTDANGDKVTTTFVKTDLVELPLQIY 2132  
Qy 1715 -----RGLVQTDNGTYFDNGYKWRNQTINAGAMIYTLDENG-KLIKASYN--SDAEXP 1767  
Db 2133 NYDANGNKIVKADGKGYELNADGTASNKEVTLG-----NYDANGKKVVKYVTEGADRWY 2188  
Qy 1768 TSTD 1771  
Db 2189 TNAD 2192  
RESULT 11  
US-09-815-242-12996

; Sequence 12996, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-03-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 12996  
; LENGTH: 6281  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; US-09-815-242-12996

Query Match 3.4%; Score 318.5; DB 10; Length 6281;  
Best Local Similarity 20.0%; Pred. No. 9.8e-10;  
Matches 395; Conservative 237; Mismatches 806; Indels 539; Gaps 96;  
Qy 12 SGQKWTAATAVAVTAVSTALLYGV--AHADQVQVQASTTQDQSTSYNNNDTKTVALDNT 69  
Db 1772 AGVQVSTVASELNTAMSLQNGINDEAATKAAQKYTDADREKQATYNDA--VTAAKTL 1829  
Qy 70 DQSAQTTDKK---QVVSNTNOSKT---DOTSTADKNSTSTPVSVPSPNTEKQAKYN 121  
Db 1830 DKTAGSNDNKAAYEQALQRVNTAKTALNGDERLNEAKNTAKQOVATM-SHLTDAQKANLT 1888  
Qy 122 EQ-----DKNGYGNIDTAYFSNNQLH--VSGWNAITNAS---OGTNS-RQLIYR 163  
Db 1889 SQTESGTTVAGVGIOQANAGTLDQAM---NQLRQSTASKDATKSEDYQDANADLQNAYN 1945  
Qy 164 DITTNELGRTDVTNN-VARPD-----VKNVHVNYRADN--SGFDVNVNIDFSKMKD 212  
Db 1946 DAYTNAE-GIISATNPENPDTINOKASQVNSAKSALNGDEKLAARAKTAKSDIGRLTD 2004  
Qy 213 YRDSIEIVSYSGNSGVDMWSQP-----ITFDKNNTAYLDTFEVKNGEL-HATG-WNATN 266  
Db 2005 LNNA---QRTAANAE--VD--QAPNLAAVTAAKNKATSLNT---AMGNLKHALLAEKDNTK 2054  
Qy 267 SATNYNHHFVILFDQTNKEVARCEVREGSRPDVAKVYPQVVGAAANSGFNVTFNISDLD 326  
Db 2055 RSVNYT-----DAQPKQOAY-----DTAVTQAEAITNANG-----SNAN 2089  
Qy 327 YTHQYQVLSRYSNSDRNG-EGDNVTVWFPQSIAPANQSQYLDSDISKNGE----- 378  
Db 2090 ETQVQAAALNOLNQAKNLNGDN-----KVAQAKESAKRALASYNLNNAQSTAATSQ 2141  
Qy 379 ----VTVTGWNA-----TDLSELQN-----NHVYILFDQTAGQVNA----- 410  
Db 2142 IDNATTVAGVTAQAQNTANELNTAMGQLQNGINDQNTVKKQVNFDTADQCKKDAYTNAYTN 2201

Qy	411	-----SAKADILISRPDVAKAYPTVKATN-----	-----SGFKYTPKVNLOP	449
Db	2202	AOGLIDKXHGQWTKAQVEAALNQVTTAKNALNGDANVQAKSDAKANGLTTLHNLNAQK	2261	
Qy	449	GHQSVYSVRSADENGNGDKRHTDYWFSVILNOTASNIDTITMSTNGLHI--AGWMAS	506	
Db	2262	QDLTSQIEGATTVNGVNGVKTRAQDLDGAMQRLQSAIANQDKTASENVIDADPTKKTAF	2321	
Qy	507	DNSTNETTPVAILLNGKEVTRQKMSL--TARPDVAAYPSL-----	556	
Db	2322	DNAITQAEY--LNDKHGANKDKQAVEQAIQSVTSTENALNGDANQRAKTEAIAIDN	2378	
Qy	557	TIKLTNDQYQALNGQLVLLRFSKAAD--GNPSGDNTVTDQFSKNYATPGGNFYVXVNG	614	
Db	2379	LTHLNTPOKTALKQOVNAAGVSGVTDLKNSATSLNNAMDOLKQAIA---DHDITIIVASG	2434	
Qy	615	NOVEFSGWHATNOSDKDSOWIIVLVNGKEVKRQLVNDTKEGAAGFENDVYKYNPALEN	674	
Db	2435	NY-----TNASPKQRAY-----	2459	
Qy	675	SSMSGFGOIT--LPVTVKENNVLVHRESNDVKTEGEGYVDFWSELPWDFSKQNGP	732	
Db	2460	---GSPNVITNAADYTAATQRY-----NNAETGLNGDTNLATAKQOAKDALRQMT-H	2507	
Qy	733	LKQFGLGTINGQYYIDPTGTQPRKNFLLOSNNWIIYFDSDTGYCT---NALELOPRAKGT	789	
Db	2508	LSDAQKOSITGO---IDSAT-----	2540	
Qy	790	VSSNEQYRNG-----NAAYSDDKSIEENVGYLTADTWYRQIILKDGTTWTDSKETD	842	
Db	2541	-NAMNOLRNSIANKDDVKASQPYVDADRDQKQAYNTAVT--NAENI-----NAT	2587	
Qy	843	MRPTLMYWWENTLTOAYYLYNKKQHGNLLPSPALPFN---ADADPRLNHYSEIVQONI	898	
Db	2588	SOPTLD---PSAVTQA--ANOVSTINKTALGAONLANKKQETTANINQLSHLNNAAQKDL	2642	
Qy	899	EKRISSETGN-----TDLRLTLMHFVFTNPMWKNK--SENYNPSGFIQOGGFLKY	946	
Db	2643	NTQVTNAPNISTVNOYKTRAEQLDQAMERLI--NGIQDKQOVKQSVNF-----	2688	
Qy	947	ENSDLTPYANSYRLLGRMPINIKOOTYRGQBFLANDIDNSNPVQAEQLNWLYYLLNF	1006	
Db	2689	--TDADPEKOTAYNNAVTAENIINO-----ANGTNAQSQVEAA-----L	2727	
Qy	1007	GTITANNQANFOSVRDAPNDIDADLMNIADQYPNAAAYGMDSDAVSKH-----IN	1058	
Db	2728	STVTTIKQALNGRKYTDKANNANOTLSTL--DNLNNAOKGAVTGNIQOHTVAEVTQAIQ	2786	
Qy	1059	ILEDWNHADPEYKNKIGNPOLTW-----DDTIKNSLNHGLSDATN-----R	1099	
Db	2787	TAQELNTAMGNLNSLNDKDTTLGSONFADADPEKKNAYNEAVHNAENILNKSTGTNVPK	2846	
Qy	1100	WGLDAIVHOSLADREN--NSTENWVPINYSFVRAHDHNSOD---QIQNAIRDVYTKOYHTF	1155	
Db	2847	DQVEAAMNQVNATKAALNGTQNL-----EKAKQHANTAIQGLSHLTNAQEAALKQLVQOS	2901	
Qy	1156	TFEDEQKG-----IDAYI-----	1191	
Db	2902	TTVAEAQGEQKANNVDAAMDKLROSADNATTKQNYTDASQNKKDAYNNAVTTAQGI	2961	
Qy	1192	TNKDITPRVYIGDLYTDGOYMEHGTTRYDYDLTLNLLSRKYVYAGGOSMOTMSGVNNNI	1251	
Db	2962	IDOTTSTPLDPTVINOAGQ-----VSTTKNALNGNENLEAAKQ--QASQSLGSLDNL	3012	
Qy	1252	LTSVRYG-----KGAMTATDTGDTRTQGGIVGVVSYNTPNLKLVNDK-----VLHMGA	1301	
Db	3013	NNAOKOTVTDQINGAHTVDEANQIKQNAONLNTANG---NLKQAIADKDATKATVNETDA	3069	
Qy	1302	AHKNOQYRAAVLTTTQDGVINTYSDOGAPVMTDENGDIYLSHNLVYNGKEEADTAVOGY	1361	
Db	3070	DQAKOQAYNTAVTNAENIISKANGNATQAEVEQ-----ALQYVNAAKQALNGN	3118	

Qy	1362	ANPDVSGYLAVWPVGASDNODARTAPSTPEKXNGNSAYR-----TNAEFDNSVIEFAF	1414
Dz	3119	AN-----VOHAKDEATALINSSNDLNOAQOKDALKOQOVNATTVAG-----	3159
Qy	1415	SNEVYTPTKESERANVRIAQNADFFASLGFTSFEMAPQYNBSKORTFLDSID--NGY--	1470
Dz	3160	-NNVKOTAQSELNAMTOLKOG-----IADKEOTKADGCFNYADPDKONAYNQ	3205
Qy	1471	AFTDRYDLGSEPNKYGTODELRNAIQ-----ALHKAGLOVMADWVPDQI	1515
Dz	3206	AVAKAEALISATPDVVVTSEITAAUNKVTOAKNDLNGNTNLATAKNQVHAIOQLN--	3263
Qy	1516	YNLPKGK-----VATVRVDGRNWKKDAIINNLIYVN--TIGGEYQKKYGG	1562
Dz	3264	LNOAQRDEYSKOITQATLVPNVNAIQQAATTLLMADANTOLKOCIANKAQIKGE---NYHD	3320
Qy	1563	AFLDKLOKLYPEIFTKQ----VSTGVADPS---OKTEWSAKVFNGNTILHRGSYYVL	1615
Dz	3321	ADTDK-OTATDVNAVTRKABELLKOTTNPMDPNTOIALUTK-----VNDTNOALNGSKLA	3374
Qy	1616	KADGGQYNYLGT-----TTKQPLFIQTGEEKKQNEGFKVGNDCGYFYFDLAGNMVK-N	1668
Dz	3375	DAQODAKTLTGTLDLNDLAKQALTIQVEQAADIATVNVVKONQON-----LNAMTNLN	3429
Qy	1669	TFIEDSVGNVYFFDOQCKMVENKHFFVDVDSYGEXGTYTFFLKNYSVFRGLGVOTDNGT	1725
Dz	3430	NALODKT-----ETLNSINFITDADO-AKKDAY---TNAYSBAIGLSUKSANG	3472

RESULT 12

```

US-09-815-242-12611
? Sequence 12611, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlson, Kari L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIORITY APPLICATION NUMBER: 60/131,078
? PRIORITY FILING DATE: 2000-03-21
? PRIORITY APPLICATION NUMBER: 60/206,848
? PRIORITY FILING DATE: 2000-05-23
? PRIORITY APPLICATION NUMBER: 60/207,727
? PRIORITY FILING DATE: 2000-05-26
? PRIORITY APPLICATION NUMBER: 60/242,578
? PRIORITY FILING DATE: 2000-10-23
? PRIORITY APPLICATION NUMBER: 60/253,625
? PRIORITY FILING DATE: 2000-11-27
? PRIORITY APPLICATION NUMBER: 60/257,931
? PRIORITY FILING DATE: 2000-12-22
? PRIORITY APPLICATION NUMBER: 60/269,308
? PRIORITY FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 12611
? LENGTH: 3158
? TYPE: PRT
? ORGANISM: staphylococcus aureus
US-09-815-242-12611

```

Query Match 3.2%; Score 304.5; DB 10; Length 3158;  
Best Local Similarity 19.8%; Pred. No. 2.6e-09;  
Matches 385; Conservative 229; Mismatches 706; Indels 621; Gaps 90;

2

Qy	42	VQASITIQDOTSIV-----	-----	NNDDTDVTA-----	-----	64		
Db	116	IQQAATLLNDAMTQKOGIAKQAQKGS	ENYEDADTDQCTAYDNAVTKAE	LLKQTINPT	175			
Qy	65	LDNTDQSAOT-----	TDKQVYSVNTNOSKTD	TDSTADKNSTSPSVSLP	109			
Db	176	MDPNTIQALTKVNDTNQALNGNOKLADAKQ	-AAKTNLGTLDLHLDNAQKQALITQVEBAP	234				
Qy	110	S-----	NNTEKQAKNYNE-----	QDKG-----	NYGNIDTA-----	YFSNNQJLHVSG--	145	
Db	235	DIATVNVNKNQALNANMTNINNALHDKET	FLNSINFTDQAKKDAYTNAVHA	BGILL	294			
Qy	146	--	WNATNSAGTNSRQIIIVRDITTNELGRTDVT	NNVARPDKVKNVYNNADNSGFDVNV	203			
Db	295	SKANGSNASQ-TEVEQAMORVNEAKQALNGND	--NVQAKDAAKQOVITNAN-----	DLN-	345			
Qy	204	NIDFSKMKDYRDSIETVSRYSNGKSVDMWS	OPITFDKNYNYLDTFFVKNGELHA--	TG	261			
Db	346	QAKDALKQOVDRAQIVANVITIKTAQDLN	QAQTLKQGIADKDTKANGNFVNDTK	405				
Qy	262	WNATNSAINVHHFVILFDQTNKGEVAR	EGVREGOSRDPVAKVPVQVYGAANS	GFNVTFN	321			
Db	406	QNAVNAVAAHEQIIL-----	SGTPNANVDPQQAQALQ--	QVTTQA	443			
Qy	322	ISLDYTHOYQVLSRYSNSDNGEDNVTYWF	NPQSIAPANOSNOGILDSFISKNGE	VTY	381			
Db	444	KGDLGNHNLQVAKDNANTAI	DOLPNL--OPOKTALKDQVSHAE	L-----	V	488		
Qy	382	TGWNVA-----	TDLSELQNNHYV--	ILFDQ-----	403			
Db	489	TGVNAIKONADALNNAWGLTKOOIQANS	QVPSQVDFTOADQDKQQAYNNAAN	QAQIANG	548			
Qy	404	-----	TAGKQVASAKADILSRDV-----	422				
Db	549	TPTPVLTPDVTQAVTMMQAKDALNGDE	KLQAQKQDALANLDTLRLNQOR	ALRNQI	608			
Qy	423	--	AKAPTVTATNSEKFTFKVNNILQPHQY	SVVSRPSADENGNGDKRHTDWFSPV-	479			
Db	609	NQAOALATVEQTKONSONYNTAMGNL	KQG--TANKDVTKASENYHDADAKQI	ATYNAVS	666			
Qy	480	-----	ILNOTAS--	NIDTITM-----	TSNGILHATGMWASDNSNETPTAITL	NGKE	525	
Db	667	QAEGII	NOTTNPTLNPDEITRALTOVTDAK	NLNGEAKLATEKQAKDAVNAMTHL	NDQAQ	726		
Qy	526	VTQRKMSLTARPDAVYPSLNSAVSG	PDITIKLT-----	NDQYQALNGQLVLLR	577			
Db	727	KQALKGIDGOSPEIATV-----	TQVKQATSLDQAMNQLSCALINDKT	QTILL-----	771			
Qy	578	FSKAADGN-----	PSGDNTVTDOFSKNYATTCGNFDY	VKVG--NOVEFSGWEATNQ-----	627			
Db	772	-----	TDGNYLNADPDQNAQYKQAVEKAE	ALLN-----	KQSGTNEVQAQVESIINEVNAA	821		
Qy	628	-----	SNDKDSOWIILVNGK--EYKRLVND	TKEGAGFNRPDYKYNPAIENSS	MGFGCI	683		
Db	822	KQALNGNDN-----	LANAKQPKQOLANIT-----	HLNDAQK-----	QSFESQ	859		
Qy	684	IT-----	LPYTVKNENYQLVHR-----	FSNDVKTGEGNYV--	DFWSELMPVKDSFKQKNGP	732		
Db	860	ITQAPLVTDTVTINQAKAALDHAME	LLRNSIADNQATLASDYDHATAQR	QNDYNAVTA	919			
Qy	733	LKQFGLQTINGQOYYIDPTIGQPRKN-	FLQSGNWNWIFDSDTGVS-----	777				
Db	920	AKNIINOTTS-----	PTMNPDEVNRATQYNNNTKV	ALDGDENIAAQAQANRL	NQLD	972		
Qy	778	--	TNA-----	LELOFARKGT-----	VSSNEQYRN-----	GNAAYSY--	DDKSIENVNGYLTAD	820
Db	973	HLNNAQKQOLOSQIAOSSDIAAVNGH	KQTAESLNTAMGNLINAIDHQA	VEQRGNFINAD	1032			
Qy	821	TWYFPROILKDGITWTD-----	SKETDM-----	RPILMWVPNTLQTA-----	Y	859		
Db	1033	T-----	DKO-----	TAYTAVNEAEAMTINKOTG	ONAQNOPEVEQA	ITKV--	QOTTLQALNGDH	1081
Qy	860	YLYNMKOHGNLLPSALPFFNADAP	LAELNHYSEIYVQOQNEKRIS	ETGNTDWRILMH-----	916			

Db	1082	NLQVAKTNAQTADALTSUNDPQKALKDQVTAATLVAVHQIEQNANT---	LNOAMHGLR	1139
Qy	917	DFTVNNPMWKNSENWFGIOGQFGLKYENSJLTPYANSYDYL---	LG	963
Db	1140	ESIODNAATRANS-----KYINEDPEQONYDQAVQAANSIINEQTATLD	1184	
Qy	964	RMFINIKDQYRCQEFLLANDIDNSNPVQABQ-LNWLYLLNFGTTTANNQANFDSVR	1022	
Db	1185	NNAINQAATVTNTTAAALGDKVLKDXDKHAKQTYSQAYL-----	NNAQKHMEDTL	1236
Qy	1023	VDAPNDIDALMIADIFYNAAYGMD-----SDAVSNKHINILEDWMHADPEYFNIGN	1076	
Db	1237	IDS-----ETTRIAVKOD-LTEAQLDQLMNTLQOSIADK-----	DATRASSAVVNAEPN	1285
Qy	1077	POLTMDDTIKN--SLNHGLSD-----ATNRWGLDAI-----	1105	
Db	1286	KKQAYDEAVQNAESITAGLNNTINKGNVSSATQAVTTSKNGLDGVERLAQDKOTAGNSL	1345	
Qy	1106	-----VHQSLADRENNSENWVIPNYSFVRADNNKSODIQNARIDVTGDKYHTFT	1156	
Db	1346	NHLQLTPAQQAQALEQINNATRDKVAEIIAQAAQNEAMKALKESIKQDPQTEASSKF	1405	
Qy	1157	FEDBQKIDAYIQDQNSTVKKNLYNIPASYAILTNDTTPRYVYGDLYTDGQYWEHQ	1216	
Db	1406	INEDQAKDAYTO-----AVQHAQDLINKTDTPTLV-----KSVIDQA	1443	
Qy	1217	TRYDYTLNLLKSRVKYVAGGOSMTMSYGGNNILTSVRYG-----KGAMTATDGTDE	1271	
Db	1444	TQAVNDAKNNLHGDK-LAQDKQRATETLNNLSLNTPQGALEQIINNAATRGEVAQKL	1502	
Qy	1272	TRTQIGIVVYSNTPNLKLVNDKYVLHMGAAHKNOQ-----YRAAVLTDDGVINTSD	1325	
Db	1503	TEAQAALNAQME--ALRNSITDQOOQTEAGSKFINEDPKQDAYQAQVAOHAKD-LINGTSN	1558	
Qy	1326	-----QCAPVAMTDENDLYLSH-----NLVYNGKEEA-DTAVQGY	1361	
Db	1559	PTLDKAQVEQLTQGVNOAKDNLHGDKLADDKQHAVTDLNQLNSLNNPQRALESQINNA	1618	
Qy	1362	ANPD-VSGYLAVWPVGCASNDQDAPTAPSTEKNKSNSAYRTNWAQDSVIEATSNFYV-	1419	
Db	1619	ATRDEVACKLA-----EAQALDQMAQLRNSITDQOOQT-----ESGSKFINE	1660	
Qy	1420	-TPTKESERANVRIAQNADFASFSLGPTSEMAPQVNSKDRITFLDSTIDNGYATFDYDL	1478	
Db	1661	DKPOKDAYQAQVQAHAK--DLINQTG-----NPTLDKSOVBQLTQ---AVTTAKD-	1704	
Qy	1479	GMSPNKYGTDEDLR-----NAIQALHAKGLQVWADWVPOIYNLP-----GREVA	1524	
Db	1705	-----NLHGDOKLARDQOQAVTTVNALPNLHQAQOALT---DAINAAPTREVAQHVQ	1755	
Qy	1525	TVTRVDRDGRVWVKDAIINNLYVYNTIGGGEYKKYGAFLDKLQKLYPETFKKQVSTG	1584	
Db	1756	TATELDHAMETLK-----NKVDQVNT-----DKAQPNTYESTADKKCAVD	1795	
Qy	1585	VAIDPSOKITTEWSAKYFNGTN	1605	
Db	1796	QALQAAESITDPT-----NGSN	1812	

RESULT 13  
US-09-815-242-5835  
; Sequence 5835, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseibeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard



Query Match 3.2% Score 297.5; DB 10; Length 2434;  
Best Local Similarity 19.1%; Pred. No. 4.8e-09;  
Matches 401; Conservative 250; Mismatches 737; Indels 711; Gaps 99;

33 GGVAHQDQVQA-----SITQDQSTVNN-----DIDKTV 63  
144 GNYVADQEKQVDSKVTNAETISGTPNATLTVDVNSAASQVNAAKTALNGDNNLRV 203  
64 ALD--TNT-----DQSAQTDDKKQWSTNOSKTDDTSTADKNSTST 103  
204 AKEHANTIDSLAOLNNAQAKLKEQVQSATTLDDGVGVKNSST-----LNTAKGLRDS 259  
104 PVSVLPSNTEKQAKNYEQDKNGYINIDTAYPSN-----NOLHVGWGNATNASQGTNSRQ 159  
260 -----IANEATIKAGQNTDASPNNRNEYDSAVTAAKAIINQ-----TSNPTMEPTIT 308  
160 IIVRIDITNNE-----LQRTDVTNNVAPDVKNVHNVNA-----DNSGFDVNV 203  
309 QVTQVTTKEQALNGARNLAQAQKT-----AKNNLNLTINNNAQKDALTRSIDGATTVAGV 355  
204 NIDSKMKDYRDSIEIVSRVSGNGKSVDMWSQPIITFDKNVAYIDTFEVKNGE-----256  
366 NQETAKATELNNAHSLQ-----NGINDE-----TQTKQKYLDAPSKSKAYDAQVNA 415  
257 -----LHATGNNAATNSAI-----NYNHFFVILFDQTNQ-----REVARQ-----290  
416 AKALLTRASQGVNDKAAVEQALQVNSKTAAL-----NGDAKLEAKAAKQTLGLTHIN 471  
291 -----EVREGQSPDVAKVYPQVVGGAANGFNVTNFSIDLTHTQVQVLSRYSNSD 341  
472 NAORTALDNEITQATNVEGNTVTKAKAQQLDGAMGQLETIRDKDITLQSQ-----NYQDAD 528  
342 NGEQDNTYWFNP-----QSIAPANOSNOGY-----LQSFEDISKNG 377  
529 DAKETAYSQAVNAAATILNTAGTNPKADVERAMQAVTQANTALNGIONLDRAKQAANT 588  
378 EVTVGNATDLSELQNNHHVILLFDQTAGKOVASAKADILSRPDVAKAYPTVKATNSGF 437  
589 AIT-----NASDLPNT-----KQKALKAQVTSAGRVSAANGVEHTAT-----625  
438 KVTEKVNLPQGHQYSSVRSFESADENGNDKRDHTDYWFSPVILNOTASNIDITITSNG 497  
526 -----ELNTAMTALKRAADRAETKASGN-----YVNADANKRQAYDEKVTAAEN- 670  
498 LHITAGMASNSINETTPYAIILNNGKEVTRQKMSLTARPDAVAVPSPLYNSAVSGFDT- 556

US-09-815-242-5835

ORGANISM: Staphylococcus aureus  
TYPE: PRF  
LENGTH: 2434  
SEQUENCE ID NOS: 14110  
SOFTWARE: Fastseq for Windows Version 4.0

Db 671 -----IVSGTPTPTLTP-ADVTNAATQVTVNAKTQNLGNHNEVAKQNA-NTAIDGLTSL 722  
QY 557 -----TIKLTNDQYQALNGQLQVLLRFSKAADGNPSGDNVTVD--- 594  
Db 723 NGPQKAKLKEQVQCATLTPNVQTVRDNQATLNTAKMG-LRDSIANEATIKAGQNTDAQ 781  
QY 595 --QFSKNYATTGG-----NFDYKVNQNGQVEFSGHATNQSDKDSQWITVLY 540  
Db 782 NKQTDYNSAVTAAKAIIGQTTSPSMNAQBINQAKQDQVTAQQALNGQENLRTAQ-----T 836  
QY 641 NGKE-----VKRQLVNDTRKGAAGFN-----ENDYKVPATENSMSGF 680  
Db 837 NAKOHLNGLSGLTDAQKDAVKRQI-----EGATHVNEVTAQONNADALNTANTNLK-NGI 890  
QY 681 QGIITLTPVTVKNENVOLVHRFNSDVKTGEGNYVD-----FWSELMVPVKFSQKNGP 732  
Db 891 QD-----QNTIKQGV-NFTDADAERKNAYTNAVTAQEOILNKAQGP 930  
QY 733 -LKQFGLQT-----INGQYIIDPTTGQPKNF--LQSGNN-----WIYFDS 772  
Db 931 NTSKDGVEALENVQRAKNEELNGQNVNAKT--TAKNALNLTINNNAQKALKSQTEG 988  
QY 773 DT-----GVGTNALELQFAKTVSSNEQYRNG-----NAAYSYDDKSIENVNGYITA 819  
Db 989 ATTVAGVQVSTASEL-----NTAKSNLQ--NGINDEAATKAAQKYTDADREKQATAYNDA 1042  
QY 820 DTWYRQKQILKDGTTWDSKETDMRPILMVWPNLTQAY-----YLYNMKQHGNNLLPSAL 875  
Db 1043 VI--AAKTLIDLTAGSNDNKAQVEALQV--NTAKTALNGDERLNEAKN-----1088  
QY 876 PFENADADPAELNHSEIVQONIEKRISTGNTDMLRMLHDFVTNNPMNKSSENVNFS 935  
Db 1089 ---TAKQVATMSHLTDAQKANLTQI-ESGTT-----VAGVQ 1122  
QY 936 GQFQGGFKYENSDLTPYANSDYRLGLRMPINIKDQYRGQEFLLAN-DIDN--SNPVV 992  
Db 1123 GQANAGTLDAQANQL-----RQSTASKDATKSSDYQDANADQONAYNDVAT 1170  
QY 993 QAEQLNWLYYLLNFGTITA-NNDQANFDSVRVDPADPNTDADLMTAQDYFNAAQYGMDSDA 1051  
Db 1171 NAE-----GIISATNNPENPDTI-----NOKASQVNSAKSALNG-DEKL 1209  
QY 1052 VSNKH-----INLEDDNHADPEYFN-----KIGNPQ--LT 1080  
Db 1210 AAKQTAKSIDIGRLDNLNNAQTAANAEDQAPNLAAVTAANKKATSLNTANGNLKHALA 1269  
QY 1081 MDDTKNSLNHGLSDATNRWGLDAIVHOSLADRENNSTENVIPNYSFVRAHDNNSQDQI 1140  
Db 1270 EKDNTRSVNYTDADQPKQQAAYDTAVTQAEAITNANGSN---ANETOVOA---ALNOL 1321  
QY 1141 QNAIRDVTGKYHTTFEDEOKGIDAY-----1167  
Db 1322 NQAKNDLNGDNKVAQAKESAKALASYSNLNNAQSTAATSQIDNATTVAGVTAQONTANE 1381  
QY 1168 -----IQDQNSTVKYNLYNIPA-----SYAILLTNKDITPRVYVYGLYTDGQY 1212  
Db 1382 LNTAMGQLONGINDQNTVKQVNETDADQKGDAYTNAVTAQGLDKAHQO-----1433  
QY 1213 MEHQRYDYDTLNLKSRVKYVAGQSQMVSQVGGNNNI---LTSVRYGKGAAMTADTGT 1269  
Db 1434 -----NMTKAQVEAALNOVTTAKNALNGDANVROAKSDAKANGLTLLHNAQ 1481  
QY 1270 DETRTQGI-GVVVSTPNLKLGVNDKVVLMHGAHKNQYRAAALTTTDDGVINTSDQA 1328  
Db 1482 QKDLTSQIEGATTVN-----GVN-----GVKTAODLDGAMQORLOSATAN-KDQ--1524  
QY 1329 PYAMTDENGDDLYLSHNLVYNGKEEADTAVOGYANPDVSGYLAVVVPVPGASNDQDA-----1384  
Db 1525 --TKASEN---YIDADPTKKTAFDNTAQAESYLNKD-----HGANKDKQAVEQA 1569  
QY 1385 -RTAPSTKNSGNSAYRTNAAFDNSNVIPEAFSNFY--TPTKESERANVTAQADFFAS 1441  
Db 1570 IOSVSTENALNGDANLQRAKTEA---IQADNLTLHTNTPOKTALKQOVNAAQRVSV-----1622



```
Qy 993 -----QAEQNLWLYLJLN-----FGTITANNDOANDPSVRVAPDNIDA 1031
Db 1475 NVTNPNMDANAINGMANQVNTTAAALNGAQNLAQAATTAKNLNLTSI-----NNAQK 1529
Qy 1032 DLNIAIDYFNAAAGMDSADVSNKHIN-----ILEDWNHADPEYFNKIGNPOLTMD 1082
Db 1530 DALFRSIDGATTVAGVQCEAKATELNNAMHSLONGINDETQYKQIQYLDAGPSKKSAY 1589
Qy 1083 DTINSLNHLGSDATNKGWLDIAVHQSLADRENNSTENVIPNVSFVRA----- 1131
Db 1590 DQAVNAAKAILTKASGQNVKAAVEQAL--QNVNSTRKALNGDAKLEAKAAAKQILGLT 1647
Qy 1132 HDNNSO-DOIQNAIRDVIG-KDVHTTFEDEQKG-----IDAVIQDONSTVKKNY--- 1179
Db 1648 THINNAQRTALDNEITQATNVEGVNTVKAKAQLDGGAMGQLETSIRKQITLQSONYQDA 1707
Qy 1180 -----LYNPASPFAILLTNKD-----TIPRVYGDLYTDGGGYMEHQTRYD 1221
Db 1708 DDAKRTAYSOAVNAAATILNKTAGGNTPKADVERAMQAVTQANTALNGIQNLERAKQAA 1767
Qy 1222 T-LTNL--LKSRYVYVAGGSMQTMVGGNNILTSVRYGKGAMTATDTG-TDETRTQGI 1277
Db 1768 TAITNASDLNTEKKEALKAQVTSAGRYSAANGVEHTATELNTAMTALKHAIKAETKAS 1827
Qy 1278 GVYVSNTPNLKLVNDKVLVHMGAAHKNQOYRAAVLTTTDDGVINYTSQGAPVAMTDENG 1337
Db 1828 GNYVNADANKRQAYDEKV-----TAAENIVSGTPTPTLTPSDVTNAATQ-VTNAKTQLNG 1881
Qy 1338 DLYLSHNLVYNGKEADTAOVGVANPDVSGYLAVVYPVGCASDQNDQARTAPSTEK----- 1392
Db 1882 -----NHNLEV-AKQNTAIDGLTSLNGPKAKLKEQVG-----QATTLPNQVTVRDNA 1930
Qy 1393 ---NSGNSAYRTNAFDSNVIFEAFSNFVYPTPKESERANVRIAQNADFFASLGFTSFEM 1449
Db 1931 QTLNTAMKGLRDSIANEATI--KADQN--YTDASQNKQTDYNNAVTA-AKAILGQTSPT 1985
Qy 1450 --APQYNSKDRTELDSTIDNGYAFTRYDYLGNSEPNKYGTDEDLR---NATQALHKAG 1503
Db 1986 MNAQEIQAQKQVTAQKQALNG-----QENLRTAQNAKQ--HLNG 2024
Qy 1504 LOVMADWVPD---QIYNLPQKEVATVTRVDRGNVWKDAIIN--NNLYVVNTIGGGEYQ 1557
Db 2025 LSDLTNAQKDAKQKQIEG--ATHVNEVTQANNADVLTAMTNLKNQIQONTI----- 2076
Qy 1558 KKYGGAFLD 1566
Db 2077 -KQGVNFTD 2084
```

## RESULT 15

```
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12713

Query Match 3.1%; Score 290; DB 10; Length 2344;
Best Local Similarity 17.0%; Pred. No. 1.3e-08;
Matches 320; Conservative 263; Mismatches 793; Indels 510; Gaps 69;

Qy 4 KKHFKLYKSGKQWYTAATV-----AVSTALLYG 33
Db 16 KTRVRLYKSGKNWVRSKGIKEIMFKIMGLPFIHSKVSQDNOSISKMTGYGLKTAIVIG 75
Qy 34 G-----VAHADQVQOQAST-----ODQTSTVNNDIDKTVALDNTDQSAQTDKKQV 82
Db 76 GAFTVNMHL-DQAPAAADAPLITSELNTOSETVGNONSTI-----DASTSTADSTSVT 128
Qy 83 SNTN--OSKTDSTADKKNSTSTPVSVLPSNTEK-QAKNYNEQDKGNVGNIDTAFFSNN 139
Db 129 KNSSVQTSNDSVTSEKSENVTSTNNSNOEKLTSTSESTSKNTTSSSDTKSVTST 188
Qy 140 QLVHSGWNTASQGTNSRQIIVRDITTNNELGRTDVTNNVARDPVK----- 186
Db 189 SSTEQPIN-TSTNOSTASNNTSOSTTPTSANLKTSTTSTAP-VKLRTPSRLAMSTFA 246
Qy 187 -----NVHNVYADN-----SGFDVNVNID 206
Db 247 SAATTALFANTITV-NKDNLKQYMTTSGNATYDOSTGVVTLTQDTTSKGAITLGRID 305
Qy 207 FSKMKDYRDSIETVRSYSGKSGVDWWSQPIITFDKNYAYLDTFEYKNGELHATGNATN 266
Db 306 SNKSPHSGKVNGLNKGNGHGGDGIQ-----FAFSPGVLCETGLNGAA 350
Qy 267 SAI-NYNHHFVILFDQTNCKEVARQEVREGSQRPDAKVPV-QVYGAANSNGVNFENISD 324
Db 351 VGIGLSNAPGFKLDYTHNTSTPNSSAK-----AKADPSNVAGGGAGAFVT----- 397
Qy 325 LDYTHQYQLSRYSNSDNGEGDNVYWFNPQSIAPANQSNQGYLDSFDISKNGEV----- 379
Db 398 ---TDSYGVASTVTSST--ADNAA---KLVNQPTNNTFQ-----DFDINYNGDKVMVT 444
Qy 380 -----TVTCW-----NATDLSLQNNHY-----VILFDQ 403
Db 445 KYAGQWTRNISDWIAKSGTTNPSLSMTASTGGATNLQOVQFGTFEYTESAVTQRYVDV 504
Qy 404 TAGKQVA-----SAKADLISRPD-----VAKAY-----PTVKATANS 435
Db 505 TTGKDIIPPKTYSGNVQVVTIDNQOASALTAKGYNVTSDSYASTYNDNTNKTVR-WINA 563
Qy 436 GPKVFP-----KVNNLPQCHQYSVISR-----PSADENGNGN-----DKR 470
Db 564 GOSVTYFTDVKAPVTVGNQTIIEVQKTMNPVLVLTDTNGTGTGTNTVTGLPSGLSYDSA 623
Qy 471 HTDYWFSPVILNOT-----ASNIDITMTSN-----G 497
Db 624 TNSIGTPTKIGQSVTVVSTVTOANNKSTTTTINNVDTTAPTPTPIGDKSSEVSPISP 683
Qy 498 LHIAGMASDINSINETTPYAILNNGKEVTROKMSLTARP-DYAAVYPSLYNSAVSGEDT 556
Db 684 INIA---TODNSGNVAVTNTVGLPSGLTFDSTNNTISGTNIGTSTITVSTDSAGNKT 740
Qy 557 TIKLTNDQYQALNGQLQVLLRFESKAADGNPSGDNTVTDQFSKNYATTGGNFDYVKVN--- 613
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 15:04:52 ; Search time 62 Seconds  
(without alignments)  
2761.539 Million cell updates/sec

Title: US-09-995-749A-2  
Perfect score: 9424  
Sequence: 1 MEIKHFKLYKSGKWVTA.....SDAEYFTSDVGRMLDQNKL 1781

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB	ID	Description
1	2483	26.3	1508	2	T31098	probable dextran
2	2446	26.0	1431	2	A45866	dextranucrase (EC
3	2373.5	25.2	1449	2	T30857	glucosyltransferas
4	2373.5	25.2	1449	2	T30352	glucosyltransferas
5	2347.5	24.9	1375	2	JT0345	dextranucrase (EC
6	2276	24.2	1475	2	B33135	gtfB protein precu
7	2274	24.1	1592	2	A38175	glucosyltransferas
8	2250	23.9	1290	2	JC5473	dextranucrase (EC
9	2191.5	23.3	1365	2	A41483	glucosyltransferas
10	2190.5	23.2	1577	2	T30858	glucosyltransferas
11	2093	22.2	1518	2	A44811	glucosyltransferas
12	2005.5	21.3	1599	2	S22737	glucosyltransferas
13	424	4.5	2817	2	B97033	uncharacterized pr
14	344.5	3.7	2334	2	S32920	cell wall-associat
15	321.5	3.4	4688	2	F82885	hypothetical prote
16	318.5	3.4	6713	2	B89921	hypothetical prote
17	317	3.4	2399	2	F71879	toxin-like outer m
18	315.5	3.3	2529	2	B64635	toxin-like outer m
19	313	3.3	2902	2	C71953	toxin-like outer m
20	311.5	3.3	1612	2	AB1347	probable peptidogl
21	311.5	3.3	4919	2	T31105	hypothetical prote
22	309	3.3	2893	2	A64556	toxin-like outer m
23	308	3.3	3194	2	D71517	toxin-like outer m
24	305.5	3.2	1806	2	AF1717	probable peptidogl
25	304.5	3.2	1463	2	T30290	AAS surface protei
26	302	3.2	1939	2	D97316	probable S-layer p
27	298.5	3.2	5005	2	F82884	hypothetical prote
28	297	3.2	2020	2	C48399	ABC-type transport
29	297	3.2	2500	2	G71609	hypothetical prote

30 293 3.1 4199 2 S76412 hypothetical prote  
31 286.5 3.0 1296 1 HMS01F aggregation protei  
32 286.5 3.0 1983 2 G86643 hypothetical prote  
33 286.5 3.0 2514 2 F81045 hemagglutinin/hemo  
34 286 4.0 3283 2 AC1018 large repetitive p  
35 283.5 3.0 2481 2 D90011 FmtB protein [impo  
36 283.5 3.0 4152 2 T31102 filamentous hemagg  
37 283 3.0 1306 2 S22624 aggregation protei  
38 283 3.0 2295 2 B71621 probable membrane  
39 282.5 3.0 2340 2 B71704 cell surface antig  
40 279.5 3.0 1999 2 AB2018 hypothetical prote  
41 278 2.9 1335 2 T30211 autolysin E - Stap  
42 278 2.9 1649 2 C86822 hypothetical prote  
43 277.5 2.9 2044 2 AB1180 probable peptidogl  
44 277 2.9 2660 2 E85822 probable invasiv  
45 276.5 2.9 2364 2 I40884 cytotoxin L - Clo

ALIGNMENTS

RESULT 1  
T31098  
Probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C;Accession: T31098  
R;Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159: 307-315, 1998  
A;Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A;Reference number: 220981; MUID:98164374; PMID:9503626  
A;Accession: T31098  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1508 <MON>  
A;Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A;Experimental source: strain NRRL B-1299  
C;Genetics:  
A;Gene: dsrB  
C;Function:  
A;Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 26.3%; Score 2483; DB 2; Length 1508;  
Best Local Similarity 41.7%; Pred. No. 1.9e-109;  
Matches 563; Conservative 188; Mismatches 438; Indels 162; Gaps 38;  
QY 477 SPVILNOTASNIDITMTNSGLHIAGWASDINSINETTPYAILNNGKEVTRQ-----K 530  
DB 28 STIMLSMTA-----TSQVNA-----DSTNVTDKSVTVSNNSNTTNGHDTWVK 72  
QY 531 MSITARPD-----VAAYPSLYNSAYSGPDTTIKLNQYQALNGQLQVLLRFSKADGN 595  
DB 73 QTIPVRKNDQTOQIAA--NATQAEKVRASDTT---TDTQQA-----ETANNT 115  
QY 586 PSGD--NTVTDQPSKNYATGNGFDYKVNQVPSGWHATNQSN-----DKDSQ 634  
DB 116 NKEDSIDNLTQKLPAYTPANQKTGYLXKDGK-----WYVTSNDNTLAKGLTTVDNHQK 169  
QY 635 WIIVLVNGKEVKQLVNDKREKGAAGFNNDYKVNPAIENSMSGFGQIITLPVTYKVEN 694  
DB 170 YFD--NNGVQAKQGFVTDNSK-----TYILDENSGN-AVTGIQQIGSQTLAF--NDN 216  
QY 695 VOLVHRFSNDVKTEGNYVDVFWSELMPVKDSF---QKNGPLKQFGLQTINGQOYIDPT 751  
DB 217 -----GEQVFADFYT--APDGKTYFYDDKGGQATI---GLKAINGHNYFD-S 257  
QY 752 TGQPRKNFLQSGNNWIFDSDTGVGTNALEQFAKGTVSSNEQYRNGNAAYSYDDKSTE 811  
DB 258 LGGLKKGFTGVIDQGVRYFDQSGQGVSTDSQIKBGLTSQTTDYTAHNAVHSTDSADF 317  
QY 812 NVNGYLTAATWYRPPKQILKDGTTWDSKFTDMRPILMWVWPNTLTQAYLYLNKQGNLL 871

Db 318 NFNGYLTAASSHYRPFKDLVRNQHEATANDFRIVSWPSPKOTQVNYLNYSOMG--L 375  
Qy 872 PSALPFFNADAPALNHYSEIVQONIEKRISGTNTDWLRTLMHDFVTNNPMNKDSEN 931  
Db 376 IDNRQMFSLKDNQANLACTVQQAETKIGVANSTAWLKTADDDIRTPQPNWMSSED 435  
Qy 932 VNFSGIQOGGFLAYENSDLTPYANSYRLLGRPIN-----IKDOTYRGGEFLLAND 984  
Db 436 P--KNHDLQNGALTIVNSPLTPDTPNSNFRLLNRTPTNOTGPKYTIIDQSGGFELLAND 493  
Qy 985 IDNSNPVQABOLNWLNYLLNFGTITANNQANFSDVRDAPDNIDADLMNIAQDYFNAA 1044  
Db 494 VDNSNPVQSEQLNHLVLMNFGSITANDSAAENFGRVDAVNDVADLLQIAADYFKAA 553  
Qy 1045 YGMD--SDAYSKNKHINLEDNHADPEYFNKIGNPOLTMDTITKNSLNHGLS--DATNHWGL 1102  
Db 554 YGVKDNATANAHLISLEDWNSHNDPEYKDPGNQNLTMDDYMHMTQLTWSJTKDMRMGTM 613  
Qy 1103 DAIVHQSILADRENNSNVIPNPSFVRAHDNNSQDQIGNAIRDV--TGKDYHTFFPEDE 1160  
Db 614 QRENDYILVNRNHDSTENTAIPNPSFVRAHDSVQTVIAQIISLHPDVXNSLAPTADQL 673  
Qy 1161 QKGDAIYQDQNSVTKYKYNLYNIPASVAILITNKDTPRIVYGDLYTDGGQYMEHQPRY 1220  
Db 674 AEAFKIYNDKQADKKYQYQNMPSAYAMLLTNKDTVPYVYGDLYTDGGQYMANKSYPF 733  
Qy 1221 DTLNLKLSRVKYYVAGGOSMOTMSVGGNNILTSVRYGKGMTATDTGTDTETRTQIGIVV 1280  
Db 734 DAINGLLKSRKYYVAGGOSMAV----DQNDILTNVRKGMSTVDSGNADTRTQIGLVI 799  
Qy 1291 VSNTPNLKLVNDKVVLMHGAHKNQYRAVLTTTGGVINYNTSDQAPVAMTDENGDLV 1340  
Db 790 VSNKENLAKSGDVTVLHMGAAHKNQAFRLLLGTAGNLSYVND--NAPKYVTNDQGDLLI 848  
Qy 1341 LSSHNLVYNGKEADTAVOGYANPDVSGYLAVWVPVGCASDNODATAPSTEKNSGSAVR 1400  
Db 849 FD-----NTEIYGVNRPQVSGFLAVWVPVGCADSHQDARTLSDDTAHDKGTFH 896  
Qy 1401 TNAAPDSNVIPEAFSNFYPTTKESERANVRIAQNAOFFASLGTPSPMAPQVNSSKDR 1460  
Db 897 SNAALDSQVIYEGFSNFQAPATNEDYTNVIAKNGQLFKDWGITSQLAPOYRSSTDTS 956  
Qy 1461 FLDSGTINDGYAFTDRYDLGMSEPNKYGTDELRNAIQALHKAQGVNADAVPQIYNLPG 1520  
Db 957 FLDSIQNGYAFTDRYDLGYCTPKYGTVDQLRAIKALHANGIAQIADWVPQIYNLPG 1016  
Qy 1521 KEVATVTRVDRGNVWKDAIINNLYVYNTIGGGEYQKKYGGAFDLKXKLYPEIFTKKQ 1580  
Db 1017 QELATVTRTNSYGRKDTNSDIDQSLYVTSRGGKGYQAQYGGAFSLDIQKKYPALFETKQ 1076  
Qy 1581 VSTGVAIDPSOKITEWSAKYFNGNILLHRCGYVLLKADG--QGYNL--GTTTKQFLPIQL 1637  
Db 1077 ISTGLPMDPSOKITEWSKIFNGNIOGKGAGYVLLKDSGTDQYKYVTSNNNRDPLPKQL 1136  
Qy 1638 TGEKKQCNQEGFVKNDGNGNYFYDLAGNMVKNFTIEDSVGNWYFDDQDGKVENKHFVDV 1697  
Db 1137 TDLSE--TGFVRDNIGMVY--YTLSSYLARTPIODDNGNYYFDSGHLVTG--FQIN 1191  
Qy 1698 SYGEGKGYFFLKNVSPRGGLVQ--TDNGTYFYFDNYGKWRNQTI-----NAGAM 1745  
Db 1192 NH----HYFFLPNGIELVQSLQWADGSTIFYDQKGRQVFNQYITDQTGYAYFQNDGTM 1247  
Qy 1746 IYT-----LD-----ENGLIKASYNSDAE 1765  
Db 1248 VTSGFTEIDGHKQVYFKNGTQVKGQFVSDT 1278

## RESULT 2

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A45866

R: Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A: Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco  
A: Reference number: A45866; MUID: 91100958; PMID: 2148600  
A: Accession: A45866  
A: Status: Preliminary  
A: Molecule type: DNA  
A: Residues: 1-1431 <HON>  
A: Cross-references: GB:M29296  
C: Superfamily: cpl repeat homology  
C: Keywords: glycosyltransferase; hexosyltransferase  
F: 181-201/Domain: cpl repeat homology <CP1>  
F: 1127-1146/Domain: cpl repeat homology <CP2>  
F: 1192-1211/Domain: cpl repeat homology <CP3>  
F: 1257-1276/Domain: cpl repeat homology <CP4>  
F: 1277-1297/Domain: cpl repeat homology <CP5>  
F: 1321-1340/Domain: cpl repeat homology <CP8>  
F: 1341-1361/Domain: cpl repeat homology <CP6>  
F: 1385-1404/Domain: cpl repeat homology <CP7>

Query Match 26.0%; Score 2446; DB 2: Length 1431;

Best Local Similarity 46.7%; Pred. No. 9.8e-108;

Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

Qy 738 LOTINGQQYIIDPTTQCPKRNFLQSGNNWIFYDSDTGVGTNALELQFAKGTVSSNEQYR 797  
Db 174 VKQIDCKYVYIG--SDQCPKKNFALTNNKVLFDKNTGALTDTISQYQPKQGLTKLNNDY 232  
Qy 798 NGNAAYSDDKSIENYVNGYLTADTWYRPKQILKDGTTTWDTSKETDMRPILVAVWNTLTQ 857  
Db 233 PHNQIVNFENTSLETIDNYVTADSWYRPKILKNGKTWTASSESDLRPLMLSWPDKQYQ 292  
Qy 858 AYLNVNMGKRG-----NLLPSALPFFNADAPALNHLHYSEIVQONIEKRISGTNTDWLR 912  
Db 293 IAYLVNMGQGLGTGEN-----YTADSSQESLNLAQAQTVQVKIETKISQTOQTQWLR 344  
Qy 913 TLMHDFVTNNPMNKDSENVPFSGIQ--PQGGFLKYNSDLTPYANSDYRLLGRPIN-- 968  
Db 345 DIINSVFTQPNWNSQTESDTSAGEKDHLOGGALLYNSDKTAYANSYRLLNRTPTSTQ 404  
Qy 969 -----IKDOTYRGGEFLLANDNSNPVQABOLNWLNYLLNFGTITANNQANFSDVRV 1023  
Db 405 GKPKYFEDNSSGGYDFLLANDIDNSNPVQABOLNWLHLMVGSIVANDPEANFDGVRV 464  
Qy 1024 DAPDNTADLMNTAQDYFNAAVQMD--SDAVSNKHINLEDNHADPEYFNKIGNPOLTMD 1082  
Db 465 DAYDNVYNADLLQASDYLKAHYGVDKSEKNAIHLISLEAWSNDPQYNKDTKGQOLPID 524  
Qy 1083 DTIKNSLNHGLS-----DATN-----RWGLDAIVHOSLADRENNSNVIPNYSFVRAHD 1133  
Db 525 NKLRSLLYALRPLEKDSAKNKEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAH 584  
Qy 1134 NNSQDQIGNAIR--DVTGK--DYHTFFTEDEKQGDIDAYIQDQNSTVKKYNYLNYPASVAILL 1191  
Db 585 SEVQTVIAKIIIAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKKYQTSNPTAYALML 644  
Qy 1192 TNKDTIPRYVYGDLYTDGGQYMEHOTRYVDTLTNLKSRVYVAGGOSMOTMSVGGNNH-- 1250  
Db 645 SNKDSITRIYVYGDLYTDGGQYMATKSPYDAIDTLKARIKAAAGQDMKIIVYEGDKSH 704  
Qy 1251 -----ILTSVRYGKGMTATDTGTDETRTQIGVYVNSNTPNLKLGVNDKVVLMHGAH 1303  
Db 705 MDWDYTGVLTSVRYGTGANEATDQGEATKQGMVITSNNPSSLKLNQNDKVIYNMGAH 764  
Qy 1304 KNOQYRAAVLTITTDGVINYTSDDQAP--VAMTDENGDLVLSHNLVYNGKEADTAVOQY 1361  
Db 765 KNOEYRPLLTTKDGLTSYTSAAAKSLYRKNTDKGELVFDASD-----IQY 812  
Qy 1362 ANPDVSGYLAVWVPVGCASDNQDAPSTPEKNSGNSAYETNAAFSDNSVIFEAFSNFYPT 1421  
Db 813 LNPQVSGYLAVWVPVGCASDNQDVRVAASKANATGOVYESSALDSQLIYEGFSNFQDVR 872  
Qy 1422 TKESERANVRIAQNAOFFASLGTFSEMAPQVNSSKDRITFLDSTIDNGYAFTRDRLGMS 1481

```
Db 873 TKSDYTNKKIAQNVQFKSGVTSFEMAPQYSSSDGSLDSIIQNGYAFEDRYDLAMS 932
Qy 1482 EPNKYGTDEDLRNAIOALHKAQVMDWPDQIYNLPKGEVATVTRVDDRGVNMKDAII 1541
Db 933 KNNKYGSGQDWINAVKALHRSIGIOIADWPDQIYNLPKGEVATVTRVNDYGEYRKDSEI 992
Qy 1542 NNNLYVNTTGGG-EYQKKYGGAPLDKLQKLYPEIFTKKQVSGVADIDPSQKITESAKY 1600
Db 993 KNTLYAANTKSGNDYQAKYGGAPLSLAARQYPSIFNRTQISNGKKIDPSEKITAANKAY 1052
Qy 1601 FNGNIIHRSGSYVLKADGG-QYNNL-GTUTKQFLPQLTGERKQKQNEGVEFGVNDGN-Y 1657
Db 1053 FNGNIIHRSGSYVLKADGG-QYNNL-GTUTKQFLPQLTGERKQKQNEGVEFGVNDGN-Y 1657
Qy 1658 FNDLAGNMVKNFTFEDSVGNWYFDDQKGVNKHFFVDVDSYGEKGYFFLKNGVSPRG 1717
Db 1107 FYSTSGYQAKNSFVODAKGNWYFDDNNGHMYGLQQLN---GE--VQYFSLNGVQLRES 1160
Qy 1718 LVQF-----DNGTYFDPNGK-----DNGTYFDPNGK-----MYRNQINAGAMTYTLDEN 1752
Db 1161 FLENADGSKNYFGLHGNRYSGYSPDNDSKWRVYFDSGVNVAVLKLTNGTQYF--DQD 1218
Qy 1753 GKLKASY 1760
Db 1219 GYQVKGAW 1226

RESULT 3
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: 220909; MUID:9512197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1
C:Genetics:
A:Gene: gtfl

Query Match 25.2%; Score 2373.5; DB 2; Length 1449;
Best Local Similarity 41.5%; Pred. No. 2.7e-104;
Matches 560; Conservative 180; Mismatches 435; Indels 175; Gaps 42;

Qy 491 ITMTSGLHIAGWASDINSINETTPYAILLNGKEVTRQKMSLT-ARPDVAAYPSLYNS 549
Db 17 VTIAVTGL-----SLGAVSAVSLGTNDGVVQADEHTDVAIPDITVDGTYSND 66
Qy 550 AVSGFD--TTIKLTNDQYQALNGQLVLLRFSKAADGNPSGDNTVTQDFSKNATTTGNN- 606
Db 67 TTAQDPTVAATND---VATDQATPTATDITDT---NIVAANAVDTVATVGTDR 119
Qy 607 ----FDYKVGNGQVPSGWHATQSKNDKQSWIILVNGKEVRQLVNDTKBAAGFN 662
Db 120 AATTNDTATNDTAVD-----TTNNNTTDT---TTVTDRAATTERRATGARRGPTGRR 171
Qy 663 NDYKVNPAIENSMGSGQGIITLPVTKNENQVLVHRESNDVKTGEGNYVDFWSELMVP 722
Db 172 --APVNGNTNANN-----VTVVNDLPA-----TNNVT----- 201
Qy 723 KDSFKGNGPLKQFLOQTINQQOYIIDPTGQPRKNFLQSGNNWYFDSPTGVTGNALE 782
Db 202 -----DGPSH---IKTINGQYVE--DDGTIRKKNYVLIRIGSGSYFNAETGELSNOKE 250
Qy 783 LQFAK--GTVSNEQYR-----NG--NRAYSYDDKSIENWNGYLTADTWPRKQILKDG 833
Db 251 YRFDKNGGTGSSADSTNTNVNGDKNAFYGTTDKDIELVDGVFTANTWYRPKEILKDGK 310
```

RESULT 4

T30552

glucosyltransferase N - Streptococcus salivarius (fragment)

```
Qy 834 TWTDSKTDMPILMVWMPNTTITQAYLYNMYKOHGNLLPSALPFFNADA--DPALNHY 891
Db 311 EWTASTENDKRLLTWMPSPKAQASLYNMYKEQGLTNTQYTSFSSQTMQDAAALE--- 367
Qy 892 EIVQOONIEKRISETGNTDKLRLTMHDFVTNNPMWKNKDSNVNFSQIQFGGFLKYENS 951
Db 368 --VQKRIEIRAREGNTDWRITKFNVTQPGWNSTSENLD--NNDHLOGGALLYNNDS 424
Qy 952 TPYANSDYRLLGRMPIN-----IKDQYRGGEFLANDIDNSNPVQAEQNLWLY 1002
Db 425 TSHANSDYRLLNRTPTSGTKENPKYKTDTSNGGFEFLANDIDNSNPVQAEQNLWLY 484
Qy 1003 LLNFGTITANNOQAFDSYRVDAPDNIDLANIAQDYENAAVGM--SDAVSNKHINILE 1061
Db 485 IMIGTITGGSDEBPDGVRVDAVNVNADLLQIASDYFKAKYGADQSQDAIKHLSILE 544
Qy 1062 DWNHADPEYFNKIGNPQLTMDDTIKNSLNHG--LSDATNRWGLDAIVHOSLADRENNSTEN 1120
Db 545 AWSHNDAYYNEDTKGAOLPMDOPMHLALVSLRLPIGRNSGVPELISNLSNDRSESGKNS 604
Qy 1121 VVIPNYSFVRADNNSQDOIQNAIRD---VTGKDYHTFTFEDEKQGTDAVIQDNSTV 1175
Db 605 KRMANYAFVRAHDSFVQSIIGIILKNEINPQSTG---NTFTLDEMKKAFETNKMRSAN 561
Qy 1176 KKYNLNIPASAYAILLNKDTIPRVYIGDLYTDGGQYMEHOTRYXDTLNLLKSRKYVA 1235
Db 662 KQYQYNIPSAVALMTHKDTVPVYIGDLYTDGGQYMEHOTRYXDTLNLLKSRKYVA 721
Qy 1236 GGSGMQTMSVG--GNNN-----ILTSRVYKGGAMTATDGTDETRVOGIGVVNTPNLK 1288
Db 722 GGQDMKVNVIYIGNTNGWDAAGVLTSVRYGTGANSASDTGTAEIRNQGMAVIVSNOPAL 781
Qy 1289 LGVNDKVLHMGAAHKNQYRAAVLTDTGVTNITSDGAP--VAMTDENGDIYLSHNL 1346
Db 782 LTSN--LTINMGAAHKNQYRAVPLLTNDGVATYLNDSANGIVKYTDGNGNLTFSANE- 838
Qy 1347 VYNGKEADTAVGCIYANPDVSGVLAWVVPVGSADNODARTAPSTKNSGNSAYRTNAAFD 1406
Db 839 -----IRGINPQVDGYLAVVVPVGSADNODARTAPSTKNSGNSAYRTNAAFD 887
Qy 1407 SNVIFAFSNTVPTKESERANVRIQAQNADEFASLGFTSFEMAPQYNSSKDRTFLDSTI 1466
Db 888 SQVYIEGFSNFQDFVQNPQSYTNKKIAENANLFSKNGITSFEFAPOYVSSDDGSLDSVI 947
Qy 1467 DNGYATDTRDYLCMSEPNKYGTDEDLRNAIOALHKAQVMDWPDQIYNLPKGEVATV 1526
Db 948 QNGYATDTRDYLCMSEPNKYGTDEDLRNAIOALHKAQVMDWPDQIYNLPKGEVATV 1007
Qy 1527 TRVDDRGVNMKDALINNLYVYVNT-IGGGEYQKKYGGAPLDKLQKLYPEIFTKKQVSTGV 1585
Db 1008 TRVNNYGETKDALIDHLSYAAKTRTFGNDYQKYGAGFLDELKRLYQIFDQVQISTGK 1067
Qy 1586 AIDPSOKITEWSAKYFNGTNIHRSGSYVLKADGGQYYNLGT--TKQFLPIQLTGKKKG 1644
Db 1068 RMTTDEKIKTQWSAKYMNGLNDRGSEYVLKNGLNGY--GTNGGKVSPL-KVVG-SNOS 1123
Qy 1645 NEGFEVKGNDG-----YFYDLAGNMVKNTEIESVGNWYFDDQKGVNKHV 1694
Db 1124 TNGDNQNGDGSKEFKRLFSVRYRN--NQYAKNAFKDNDGNVYIFDNGSGRMAVGEKTI 1182
Qy 1695 DVDSYGEKGYFFLKNGVSPRGGLVQTDNG--TYIFDNYGKM-----YR 1736
Db 1183 D----GKQ--YFFLANGVQLRDGYRQNRGQVYFYDQNGVLNANGKQDPKPDNNNASGR 1236
Qy 1737 NQINAGAMIT--LDENGKLIKASYNDAE 1765
Db 1237 NQFVQIGNWYAYDNGKRVGTQHQINGQ 1266

RESULT 4
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
```

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius Vi477 gtfN.  
A:Reference number: Z20854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAP>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gtfN

Query Match	25.2%	Score 2373.5;	DB 2;	Length 1449;
Best Local Similarity	41.6%	Pred. No. 2.7e-104;		
Matches 561;	Conservative 178;	Mismatches 436;	Indels 175;	Gaps 42;

[illegible][illegible]

A: Accession: C33135  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-349 <SHI>  
A: Cross-references: GB:M17361  
C: Genetics:  
A: Gene: gtfC  
C: Function:  
C: Description: catalyzes the synthesis of both water-soluble and water-insol  
C: Superfamily: cpl repeat homology  
C: Keywords: duplication; glucosyltransferase; hexosyltransferase  
F: 1-34/Domain: signal sequence #status predicted <SIG>  
F: 35-1375/product: glucosyltransferase #status predicted <MAT>  
F: 1126-1145/Domain: cpl repeat homology <CP1>  
F: 1253-1272/Domain: cpl repeat homology <CP2>



F;1318-1337/Domain: cpl repeat homology &lt;CP3&gt;

```

Query Match      24.9%: Score 2347.5; DB 2; Length 1375;
Best Local similarity 41.3%: Pred. No. 4.2e-103;
Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;

QY 549 SAYSGFDTTIKLTNDQYOALNGQOLVLRFSKAADG--NPSGDMNTVTDQFSKNVATTGNN 606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 TSLSGSILVRADSTDRQQAIVTSSQASLVTSEAAKEILTATDSTATSAISQPTAIVTDN 88

QY 607 FDYKVGNGQVFSGHWATNQSNDRKDSOWI IVLVNGKEVKQLVNDTKEGAAGFNNDVY 666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 VS-----TTNQS-----TNTANTA-----NFVV 107

QY 667 KVNPATENSSMSGFOGIIPLTVYVAKENVQLVHRFSNDVKTGECNVY-----DEW 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 KPITTSQAKTDNSDKIIITSAVNR-----LTATSKFVPANNTAHPKVT 154

QY 717 SELMPVKDFSKNGPLKQ-----FGLQTINGQOYYIDPTTGQPRNFLL 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 DKIVP IKPI---GKLKQPSLSQDDIAALGNVKNRKVNGKYIY- KEDGTGLQNYAL 209

QY 762 QSGNNNIYFSDTGVGTNALELQFAKGVTSNE---QVRNGNAAYSYDDKASIEWNGYLT 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 NINGKTFEDETGALSNNLTSPK--KGNITNNDNTNFAQVNVYSTDVANFEHVDHYLT 267

QY 819 ADTWYRPKQLDKGTWTDTSKETDRPPIIAWVWNTLTQAYLYNWKQHNLPLPSALPFF 878
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 AESWYRPKYLLDKGKTWTOSTEKDRPPLMTWWPDEQEQOYVYNNMAQLGIHOT---Y 323

QY 879 NADADPAELNHYSEIVQOQNIERRISGETNDLRLTLMHDFVNTNPMNKKDSENVNFSG 938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 NTATSPQLNLAAOTIQTKEEKTAEKNTNWLHOTISAFVKYTSAWNDSDEKPFDDHLQ 383

QY 939 FOGGLFKYBNSDLTPVANSDYRLLRGMPIN-----IKDQYRQOEYLLANDIDNSN 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 -KCALLYSNNKLTQSANSYRIILARTPTNQTGKKDRPYADRTIGGYEYLLANDVDNSN 442

QY 990 PVYQAEQLNLWYLLNFGPITANDQANFDSVRVYDAPDNIDADLMNTAQDYFNAAYGM-D 1048
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 PVYQAEQLNLWHLFNMFGNIYANDPANFDSIRVDVANDVDADLLQAGDYLKAAXGINK 502

QY 1049 SDAVSNKHINILEDWNHADPEYFNKIGNPQLTMDDTTKNSLNHGLSDATN- RGLGLAIVH 1107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 NDKAANDHSLISLEAWSYNDTPYLDHDDGNNMNDNRULSLLYSIAKLPLNQREGMPLIT 562

QY 1108 QSLADRENNSTENVIPNYSFVRAHDNNSQDQIGNAIR-----DVTGKDYHTTFFDEQK 1162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 NSLVNRTDQNAETAAPVSPYSFIRAHDSVEQDLIRNIIRTEINPNVVG---YSFTTEIKK 619

QY 1163 GIDAYIQDQNSTVKKYLNYNIPASVAILLTKNDITPRVYTGDLTGDGQYMEHQTRYDT 1222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 AFEYKNDLLATEKYTHYNTALSVALLLTKNSVPRVYTGDMETDDGQYMAKRTINYEA 679

QY 1223 LTNLKSRYKVYAGGSGMOTMSVGGNNNLTLSRYGKGAMTATDGTDETRTQIGVYVS 1282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 680 IETLLKARIYVSGGQAMENQOV--GNSIIITSVRYGKALKATDGTDRTRTTSQGVAVIEG 738

QY 1283 NTPNLKLVNDKVLVHMGAAHKNQOYRAVLTITTDGVINTVTSQDGP--VAMTDENGDIY 1340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 739 NFPSLKLASDRVYVNGNAHKNQAYRPLLTDTONGIKAHYSQOEAGLVRYTNDRGELI 798

QY 1341 LSSHNLVNGKBEADTAQGVYANDPVSGYLAWVPVQASNDQARTAPSTKXSGNSAYR 1400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 799 FTA-----AD--IKGANPOVSGYLGWVPVQAAADQDQVRYAASTAPSTDGSKVH 846

QY 1401 TNAAFDSNVIFEAFSNVYVPTTKESERANVRIAQNAPFEASLGFTSFEMAPQVNSSKDT 1460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 847 QNALDSRVWFEGFSNFQAFATKKEEYTNVVIANKVDFAEWGVTFDEMAPQVSVSTDSG 906

QY 1461 FLDSITDNGYATDRYDLCMSBPNKYGPDRLNAIQALHKAHGLQWMDWVPDQIYNLPGE 1520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 907 FLDSYIQNGYATDRYDGLIGSKPNKYGPADDLVRAIKALHKSGLTKYNADWVPQOMTALPE 966

```

Qy	1521	KEVATVTRVDRGNVKKDALINNNLTVYN-TGGGEYKQKKYGAFLDKLQKLYPBIFTKK	1579
Db	967	KEVYATVTRVDRGNVKKDALINNNLTVYN-TGGGEYKQKKYGAFLDKLQKLYPBIFTKK	1026
Qy	1580	QVSTGVADPSQKITEWSAKYFNGTGNILHRGSGYVLKADGGGYVNLGTTTKQFLPIQLTG	1639
Db	1027	QISTGVDPDPSVKLKQWSAKYFNGTGNILHRGSGYVLKADGGGYVNLGTTTKQFLPIQLTG	1085
Qy	1640	EKKOGNSEGTVKG--NDG-NYYFYDLAGNMYKNTFTEDSVG-NWYFFDQDGKAVKNKHVFD	1695
Db	1086	NPNHGTSSSYTGLVFDGKGYVYSTSGNAKNAFI--SLGNWYFYDNGGYVAVTGAQGIN	1143
Qy	1696	VDSYGEKGTFFFLKNGVSPFRGGLVQT-----DNGYV-----YFDNGVK	1733
Db	1144	-----GANYFYLSNGIOLRNIAYDNGNKNVLSYGNDRRYENGYLFGQGWRYFN-GI	1196
Qy	1734	MVRNQTINAGAMTYTLDENKLIKASYNSDAEYPTSTDVGM	1775
Db	1197	MAVLGTRVHGAVOY-FDASG-----FOAKGQFITAD-GKL	1230
RESULT 6			
B33135			
glfB protein precursor - Streptococcus mutans			
C:Species: Streptococcus mutans			
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999			
C:Accession: B33135; A:33128			
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			
J. Bacteriol. 169, 4263-4270, 1987			
A:Title: Sequence analysis of the glfB gene from Streptococcus mutans.			
A:Reference number: A33135; MUID:87308013; PMID:3040585			
A:Accession: B33135			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1475 <SH1>			
A:Cross-references: GB:ML7361; NID:q153639; PIDN:AAA8588.1; PID:gl53640			
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			
submitted to the Protein Sequence Database, September 1990			
A:Reference number: A33128			
A:Accession: A33128			
A:Status: preliminary; not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-171,173-641,'N',643-1475 <SH2>			
A:Experimental source: strain GS-5			
C:Superfamily: cpl repeat homology			
F:1096-1115/Domain: cpl repeat homology <CP1>			
F:1244-1243/Domain: cpl repeat homology <CP2>			
F:1289-1308/Domain: cpl repeat homology <CP3>			
F:1354-1373/Domain: cpl repeat homology <CP4>			
F:1419-1438/Domain: cpl repeat homology <CP5>			
Query Match 24.2%; Score 2276; DB 2; Length 1475;			
Best Local Similarity 42.4%; Pred. No. 1,1e-99;			
Matches 518; Conservative 193; Mismatches 382; Indels 140; Gaps			
Qy	637	IVLVNGKEVARQLVND-----TKGAAGFNERNDYKVNPAIENSSM	677
Db	35	LVKADSNESKQISDNDSTSVWTANEESNVITEATSKQAASSQN--HTVTTSSSTSV	92
Qy	678	SGFGIITLPTVTKNENVQLVHRFSNDVKTGEGNYVDFWSEL-----MPVKD	724
Db	93	VNPKEVSNSEYTV-GETASNGEKLQKQTTT-----VDKTSEAAANNISKQTEADTDVID	146
Qy	725	SFQKNGPL--KQFGLQTINGQQYYIDPTGQPRKNFLQSGNNWYFD-----SDTGVG	777
Db	147	DSNAANLIQLEKLPNKVEIDGKYIYD--NNGKVRTNFTLIADGKILHFDGTGAVDTSD	205
Qy	778	TNALELOFAGKTSSNEQVINGNAAVSDDKSIENVNGVLTADTWYRKQILKDGTTWD	837
Db	206	TVNKDIV-----TTRSN-LYKKNQVYDRAQSFEHVDYLTAEWYRKYILKDGKTTWQ	260
Qy	838	SKETDMPILNWWNPNTLTQAYLYNTWKQHGNNLLPSALPFFFNADAPAEALNHYSEIVQON	897

Db 261 STEKDFRPLMTWPDQETQRQYVYVYNAQIGNKTI-----YDTSNQLQNIAAATIAQK 316  
QY 998 IEKRISGTGNTDWLRLTHMDHFTVNTNPMWKNKSDSENVNFGSQIFQGGFLKYEN-SDLAPYAN 956  
Db 317 IEAKITTKNTDWLRLQRTISAFVKTSQSAWNSDSEK-PPDD-HLQNGAVLYDNECKGLPYAN 374  
QY 957 SDYELLGRMPIN-----IKDOTYRGQEFLLANDIDNSNPVQAEQLNWLYYLLNFG 1007  
Db 375 SNYKILARTPTNQKKDPKYATADNTGGYEFLLANDVDSNPVQAEQLNWLHFLNFG 434  
QY 1008 TITANNQANFSDVRDAPONIDADIAMNIADQYFNAYGM-DSDAVSNKHINILEDNHHA 1066  
Db 435 NIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHKNDKRAANDHLSILEAWSN 494  
QY 1067 DPEYFNKIGNPOLTMDTIRKNSLNHGLSDATN-RWGLDAIVHQSADRENNSTENVVVPN 1125  
Db 495 DTPYLHDDGNKNIIMNKRLSLLSLAKPLNQRSGNPLITNSLRNRTDQNAETAAVPS 554  
QY 1126 YSFVRAHDNNSQDQIQNAIR-----DVTGKDYHTFTPEDEQKGDIDAYIQDNSTVKYKYL 1180  
Db 555 YSFTRAHDSEVQDLADIIRAEINPNVVG--YSFTMEEIKKAFEIYKNDLLATERKKYTH 611  
QY 1181 YNIPASVAILLTNKDTPRYVYGGDLYTDGGQYMEHQTRYDYLTLNLLKSRVKYVAGQSM 1240  
Db 612 YNTALSALLTNKSSVPRVYVGMFTDDGGYMAHKTINYEAETELLKARIKYVSGQAM 671  
QY 1241 QTNVGGNNLLTSVRYKKGAMTATDTGTDTRTQGTGIGVYVSNTPNLKLVNDKVVYLMHG 1300  
Db 672 RNQOV-GNSEIITSVRYKKGALKATDGTDRTRTSGAVIEGNNPSLRKASDRVVVNMG 730  
QY 1301 AAHNQOYRAAVLTDTGTVNYTSDQAP--VAMTDENGDLYLSSHNLVYNGKEADTAV 1358  
Db 731 AAHNQOYRPLLLTDTGNGIKAYHSDGAAAGLVRYTNDGELIFTA-----AD--I 778  
QY 1359 QGYANPDVSGYLAVWVPVGAS--DNQDARTAPSTEKNSGNSAVRTNAAPDSNVIFEFASN 1416  
Db 779 KGYANPDVSGYLVWVPVGAALIKWALRLA--RPQOMASVHQNAALDSRVWPEGFSN 835  
QY 1417 FVYPTTESERANVRIQONADFFASLGFTSFEMAPQVNSKDRFTDSTIDNGYAFDTRY 1476  
Db 836 FQAFATKKEEYTNVVIKNDVKFAEWGVTDFEMAPQVNSSTGDSFLDSVQNGYAFDTRY 895  
QY 1477 DLGSEFNKYGTDLRLNATCALHAGLVQVADWVPQIYNLQKKEVATVTRVDDRGNVW 1536  
Db 896 DLGSKFNKYGTADLLKALKHSGIKVWADWVPQMAFPEKEVVTATRVDKYGTVP 955  
QY 1537 KDAIINNLYVN-TIGGGEYQKYGAFDLKLOKLYPEIFTRKQVSTGVAIDPSQKITE 1595  
Db 956 AGSQIKNTLYVVDGKSSGKQQAQKYGAFLEELQAKYPELFARKQISTGVMPDPSVKIKQ 1015  
QY 1596 WSARYFNGTILHRGSGYVLKADG-GQYXNLGITTK-QFLPIOLTGCKKQGNBGFVKGN 1653  
Db 1016 WSARYFNGTILHRGAGYVLKQATNTYFNISDNKEINFLPKLLNQDSQ--VGFSYDGG 1073  
QY 1654 GNYFYFPLAGNMVKNFTFIEDSVGNWYFFDDGDKMVENKHFVDVDSGEKGTFFLKNGVS 1713  
Db 1074 G-VYVYSTGSQYAKNTFISEG-DKWYFYDNGNMYMTQAQSN-----GVNYFYFLSNGLQ 1125  
QY 1714 FRGLVQTDNGTY-YFQNYGKWRN-----QTINAGAM-----IYFLDEN 1752  
Db 1126 LRDAIKNEBDGTAYAYGNDGRRYENGYYQPMGVRHFNNGNSVGLTVIDGQGVYFDEM 1185  
QY 1753 GKLIKASYNDAEYPTSTDVGKM 1775  
Db 1186 G-----YQAKGKPVTTAD-GKI 1201

RESULT 7  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C;Species: Streptococcus sobrinus  
C;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999

C;Accession: A38175  
R;Abo, H.; Mtsumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ  
A;Reference number: A38175; MUID:91123227; PMID:1704006  
A;Accession: A38175  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1592 <Abo>  
A;Cross-references: GB:D90213; NID:g217032; PIDR:BAA14241.1; PID:d1014946; PID:g21703  
C;Superfamily: cpl repeat homology  
F;1093-1112/Domain: cpl repeat homology <CP1>  
F;1122-1241/Domain: cpl repeat homology <CP2>  
F;11287-1306/Domain: cpl repeat homology <CP3>  
F;11330-1351/Domain: cpl repeat homology <CP4>  
F;1352-1371/Domain: cpl repeat homology <CP5>  
F;1402-1420/Domain: cpl repeat homology <CP6>  
F;1465-1484/Domain: cpl repeat homology <CP7>  
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 24.1%; Score 2274; DB 2; Length 1592;

Best Local Similarity 45.9%; Pred No. 1.6e-99;

Matches 496; Conservative 167; Mismatches 340; Indels 78; Gaps 29;

QY 688 VTVKNEVOLVHRFSNDVKTEGNGYVDFWSELMFVKDSFQKNGPLKQFGLQTINGQOYY 747  
Db 116 VPTENEN---QCFTDEMLAEAKNVATAESDSIP-----SDLAKSNVQVQDKYY 163  
QY 748 IDPTTQPKRNFLLQSGNNWIFVDSDTGV--CTNALEQFAKGTVSSNEQ-YRNGNAAYS 804  
Db 164 YD-QDGNVKNFAVSGDKIYIFD-ETGAYKDTSKVDADKSSAVSQNATFPAANNRAYS 221  
QY 805 YDKSINENGYLTADTYWRPKQILKDGTTWDSKETDMRPILMYWPNTLTQAYLYNM 864  
Db 222 TGAKNPEAVDNLTDADSWYRPSILKDGKTWESGKDDFRLPMAWMPDETETKRNVNYM 281  
QY 865 KQHGNLPLSALFFENADAPAEINHYSETVOQNIKRISETGNTDWLRLTHMDHFTVNM 924  
Db 282 ----NKVQIDTYTAETSQADLTAAAEVLQARIQKTISETNNTKWLREISAFVKTOFQ 337  
QY 925 WKKDSNVNFSIQGGFLKYEN-SDLTPYANSYRLLGRMPINIKQO-----TYR--- 975  
Db 338 WNGSEKPYDD--HLQNGALLFDNQTDLTPQCSNYRLLNRTPTNQTGSLDSRETYNPD 395  
QY 976 ---GQEFLLANDIDNSNPVQAEQLNWLYYLLNFGTITANNQANFSDVRDAPONIDAD 1032  
Db 396 PLGGYDFLLANDVDSNPVQAEQLNWLHFLNFGSIYANDADANFDSIRVDAEDNVDAD 455  
QY 1033 LKNIAQDYFNAAAYGMD-SDAVSNKHINILEDNHNDADPEYFNK1GNPQLTMDDTIKNSLNH 1091  
Db 456 QLISSDYLKAAYGIDKNNKNANNHVSIVEASDNDTPYLHDDGNLMMNDKFRLSMLW 515  
QY 1092 GLSDATN-RWGLDAIVHQSADRENNSTENVVVPKSVFRAHDNNSQDQIQNAIR-DVTG 1149  
Db 516 SLAKPTDVRSGNPLIHSNLSVDREVDRETVTPVSIFARAHDSQVDIRDIIRKAEINP 575  
QY 1150 KQY-HTFTEDEQKGDIDAYIQDNSTVKRYNLYNPASVAILLTNKDTPRYVYGDLYTD 1208  
Db 576 NSFQYSFTQEEIDQAFKIYNEDLKKSKDKYTHYVNPVPLSYTLTLTKNGSTPRVYGDMDTD 635  
QY 1209 GGOYMEHQTRYDYLTLNLLKSRVKYVAGQSMQVSGNNILTSVRYKGMATPDITG 1268  
Db 636 DGQYMAKNTVNYDAIESLLKARKYVAGQAMQNYQI--GNGEILT SVRYGKALKQSDKG 694  
QY 1269 TDETQGIQGVVSTPTNLKLVGNDKVV-LHMGAHKQOYRAAVLTDTGTVNYTSDQ 1327  
Db 695 DATTRTSGVGVVNGQPNFSL--DGKVVALNMGAAHQAQYKALMVSTGQVATVATDAD 752  
QY 1328 AP-----VAMTDENGDLYLSSHNLVYNGKEADTAVQGYANPDVSGYLAVWVPVGSNDQD 1383  
Db 753 ASKAGLVKRTDENGLYFLNDL-----KGVANPDVSGFLQVWVPVGAADDQD 800  
QY 1384 ARTAPSTEKNSGNSAVRTNAAPDSNVIFEFASNFTVPTTKESERANVRIQONADFFASLG 1443

Db 801 IRVAASASTDGKSLHQDAAMSGRMFEGFSFQSFATKEEBEYTVVIVANNVDKFFVSWG 860  
Qy 1444 FTSEMAPQYNSSKDRFTLSDINDGYAFTRDYDLGMEPNKXTGDBDLRNAIQALHAKG 1503  
Db 861 ITDEMAPQVYSSDQGFQDSVQNGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKG 920  
Qy 1504 LOVMADVPQIYINLPKEVATYTRVDDRGVWKDAIINNLYVWT-IGGGEYQKYYG 1562  
Db 921 LKNMADWWPDMYTFPKQEVVYTRTDKFKPIAGSQINHSLYTDTKSGDGYQAKYGG 980  
Qy 1563 AFLDKLQKYPEIFTKQVSTGVAIDPSQKITEWSAKYFNGTNIHRGSGYVLKAD-GQG 1621  
Db 981 AFLDELKEXYPELTKQMTGQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSQVSNK 1040  
Qy 1622 YNLTGTTKQFLPQLTGERKQKNEGFKVKGNDGNYFYD-LAGNVKNTIFIDSQVNWY 1679  
Db 1041 YFNVASPT-LFLPSLLGKVVESGIRY---DGKGYIYNSATGDQVKASFITEA-GNLY 1094  
Qy 1680 FFDQDGRMVENKHFVDVDSYGEKTYTFFLKNVGSFRLGYOTD-NGTYVYFDNKGKVMRN 1737  
Db 1095 YFGDGYMTGAQFIN-----GANYFFLENGTALR-NTIYTDAGNSHYIYANDGRKYEN 1147  
Qy 1738 Q 1738  
Db 1148 E 1148  
RESULT 8  
JC5473  
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C/Species: Leuconostoc mesenteroides  
C/Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C/Accession: JC5473  
R/Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996  
A/Title: Cloning and sequencing of a gene coding for a novel dextransucrase from Leucon  
A/Reference number: JC5473; MUID:97136686; PMID:8962063  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-1290 <MON>  
A/Cross-references: GB:U38181  
C/Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont  
C/Genetics:  
A/Genes: dsrA  
C/Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>  
Query Match 23.9%; Score 2250; DB 2; Length 1290;  
Best Local Similarity 46.4%; Pred. No. 1.6e-98;  
Matches 470; Conservative 155; Mismatches 336; Indels 52; Gaps 20;  
Qy 741 INGOQYIDPTTGPRKNFLQSGNNWYFDSGTGVTGNALELOQFAKGTYSNNEQYANGN 800  
Db 7 VDGKVYFVG-DDGPKPKNFTHIDGKPYFEDKRGALSNN-DKQYVSELSIGNKH--N 61  
Qy 801 AAYSVDKSIENVNGYLTATWYRKPILKDKTWTDSKTDMDRPILMWWPNTLQAYY 860  
Db 62 AVINISDNFTQLRGHTLTSWYRKPDKILANGKRWAPSTVTDPRPLMAMWPKSTQVY 121  
Qy 861 LNYMKQGNLLPSALPFENADAPAEALNHYSEIYQOQNIKRISITGNTDMLRILMHDFVT 920  
Db 122 LNYMKDQG--LLSGTHFSDNENKRTLTAAAMQAVNIKKIGQLGNTDMLKMTAMTQYID 179  
Qy 921 NNPWNKDSNVNFGSGTQFGGFLKYNSDLTPYANSDYLLGRMPINIKDQ---TYR-- 975  
Db 180 AQPWNWIDSEAKGDD--HLQSGALLYNSDMSPKANSYRKLSTPKNKGQADKYKQG 237  
Qy 976 QOPELLANDIDNSVPVQAEQLNWLNYLLNFTGNTANNDQANFDSVRVDPADNIDALMN 1035  
Db 238 GFELLANDVDNSVPVQAEQLNWLHMYMMNIGSLQNDQANFDCYRVDVNDVADLLQ 297

Qy 1036 IQADYFNAAYGMDSDAVSNKHNILIEDWNHADPE---YFNKIGNPQLTMDDTIKNSLHNG 1092  
Db 298 IAGEYAKAAYGVDK--MTRERINITYQFWKTGEMKIQTMSKHMATSKLSMDFPLHLAIKYA 355  
Qy 1093 LSDATN-RWGLDAIVHQSLADRENNSNTENVIPNYSFEVAHDNNSODQIQONAIRO---- 1146  
Db 356 LNPNDKRGSLPTEHRSLSVKRITDDKENVAQPNYSFIRAHDESVQVTIADIILKIKINPA 415  
Qy 1147 VTGKDYHTTFDEQKGDIDAYIQDQNSTYKYNLXNIPASYAILLTNKDTPRIVYGDLY 1206  
Db 416 STGLD-STVTLQIKQAFDIYNADLKADKVTPYNIPASYALLPNKDTIPRIVYGMDF 474  
Qy 1207 TDGGQYMEQTRYDYTLTNLLASRVKYVAGSGMQTMSVGGNNILTSVRYKKGAMTAD 1266  
Db 475 TDGQYMAQSPYQYDAIDALLAKRYAAGQGTMMKNVPPDEQSVMTSVRYKKGAMTASD 534  
Qy 1267 TGTDETRTQIGIGYVYSNTSNLKLGVNDKVLVHMGAAHKNQYRAAVLTITDGVINYTSQ 1326  
Db 535 SGNQETRYOGIGLVNRRPDLXLDKDEKMDGAAHKNQDYRPVLLTTKSLKYVSTDA 594  
Qy 1327 GAPVAMTDGNDLYLSSHLNVNGKEADTAVQGYANPDVSGYLAWVPVGSADNODART 1386  
Db 595 NAPVVRTDANGQLTF-----KAD-MVYGVNDPQVSGYIAAVPVGASENQDART 642  
Qy 1387 APSTEKNSGNSAYRTNAAFDNSVFIPEAFNSFYVTPTKESERANVRIAQNADEFFASLGFTS 1446  
Db 643 KSETTOSTDGSYHNSAALDSQVIYEGFSNFODPTTDPDEFNIIKAQWNLFKDWGITS 702  
Qy 1447 FEMAPQYSSKORTFLDSTIDNGYAFTRDYDLGMEPNKXTGDBDLRNAIQALHAKGLOV 1506  
Db 703 FEMAPQYRASDSKFLDAIVQNGYAFTRDYDGYNTPTKYGTADNLLDALRALHGGQIQ 762  
Qy 1507 MADWVPDQIYNLPKGEVATYTRVDDRGVWKDAIINNLYVVTITGGGEYQKYYGAFLD 1566  
Db 763 INDWVPDQIYNLPDEQLVTAIRTDGSGDGYTGSVIDHLYASKTVAGGIYQOYGGAFLE 822  
Qy 1567 KLOKLYPEITFKKQVSTGVAIDPSQKITEWSAKYFNGTNIHRGSGYVLKADGQYVNLG 1626  
Db 823 QLKTYQFLFOQKQISTDQPMNPDIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQOYFVY 882  
Qy 1627 TTTKQFLPQLTGEKQKQNEGTV-KGNDGNYFYDLAGNMVKNFTFEDSGVNWYFFDQDG 1685  
Db 883 SDAQTEFLPKQLIGEK--AKTGFTVTRGKETS--FYSTSGYQAKSAFICDN-GWNYFFDDKG 937  
Qy 1686 KMVENKHFVDVDSYGEKG-TYFFLKNVGSFRLGYOTDNGTYTTFDNYGKVMRN 1737  
Db 938 KMVVGWQVIN-----GINVYFLPNGLIQLDQAYLVHDMGYMYNNIGKQLHN 983  
RESULT 9  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus  
C/Species: Streptococcus sobrinus  
C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C/Accession: A41483  
R/Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A/Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltr  
A/Reference number: A41483  
A/Accession: A41483  
A/Molecule type: DNA  
A/Residues: 1-1365 <GIL>  
A/Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653  
C/Genetics:  
A/Genes: gtfs  
C/Supfamily: cpl repeat homology  
C/Keywords: glycosyltransferase; hexosyltransferase  
Query Match 23.3%; Score 2191.5; DB 2; Length 1365;  
Best Local Similarity 40.9%; Pred. No. 9.9e-96;  
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34;



```
QY 946 YENSOLTPYANSYRLLGRMPINIK-----DOTYRQGBELLANDIDSNPNVVAEQQLN 998
Db 473 FLNSATSASHANSRLLRNPFTNOTGRKYHIDRSNGGYELLANDIDSNPNVVAEQQLN 532
QY 999 WLYYLLNFGTITANDQANFDSVDPADPNIDADLMIAQDYFNAAFGM--DSDAVSNKHI 1057
Db 533 WLHYIMNIGSILGNDPSANFDGVRIDAVNDVADLLOIASDYKPEKRYRVADNEANAIAHL 592
QY 1058 NILEDWHADEPYENKIGNOLTWDDIKNK--LNHGLSDATNRWGLDAIVHQSLADRENN 1116
Db 593 SILSAWSYNDHQYNKDTKGAGLSIDNPLRETLTTLTKRKNYRGLSERVITNSLNRSSE 652
QY 1117 STENWVTPNKSFVRAHNNDSODQONAI--RDVTGK--DYHTFTPEDEOKGDVADVQDQNST 1174
Db 653 QKHPRDANYIFVRHDSVQAVLANIISKOINPKTGTFTFWDLQKQAFIYNADIKA 712
QY 1175 VKKNLYNPASAYAILLTNKTIDPRVYGYDLYTDGGQYMEHQTRYYDTLTNLKSRVYV 1234
Db 713 DKYITQYINPAAYATMLNDSITRVYGLDFTDDGGYMAEKSPYNAIDALLRARIKYV 772
QY 1235 AGGQSMQTMSSVGGNNILTSVRYGKGAMTADTGTDETRTQIGVYVSNTPNKLGVNDK 1294
Db 773 AGGDMKVTKLNG--YEIMSSVRYCKGAEEANQLGTATRNQGMVLVTANRPDKMLGANDR 831
QY 1295 VYLMHGAHKNQYRAAVLITTDGVINVTSDQAP---VAMTDENGDLYLSSHNLYVNGK 1351
Db 832 LVNWMGAHKNQAYRPLLLSKSTGLATYLDSDVPAGLVRYTDNQNLTTTADD----- 885
QY 1352 BEADTAQGVANPDVPGYLVAVWPVGASDNODATPTSTEKNSGNSAYRTNAAFDSNVIF 1411
Db 886 -----IAGHSTVEVSGYLVAVWPVGASENODARTKASTK--KGEQVPESSAALDSQVY 939
QY 1412 EAFSN---EYVTPKESERNVRIAQNAFPAISLGTSFEMAPQYNSKDKORTFLDSTIDN 1468
Db 939 EGFNFQDEKVTPTQYTNRV---IAQNAKLFKEWGITSFEPAPQYVSSQDGTFLDSTIEN 995
QY 1469 GYATDRYDLGMSPNKYGTDEDLNAIQALHKAQGLVADWYVDQIYNLPKGEVAVTR 1528
Db 996 GYAFEDRYDIAMKNKYGSLKDLMDLALRALHAEIGISAIDWYVDQIYNLPKGEVAVTR 1055
QY 1529 VDDRGNVWVKDAIINNLYVNT--IGGEYQKYGGAFLDKLQKLYPEIFTKKQVSTGVAI 1587
Db 1056 TNSYGTTPRPAEYNSLYAAKTRTFGDFQKYGGAFLDELKAKYPAIFERVQISNGRKL 1115
QY 1588 DPSOKITWSAKYFNGTNILHRGSGYVLKADG--GOYVNL--GITTKOFLPIQLTGEKKQ 1644
Db 1116 TTNEKITQWSAKYFNGSNIQGTARYVLQDNATNQYFSVRAGQF---FLPKQMT---EIT 1169
QY 1645 NEGFVK--GNDGNYFYDLAGNWKNTFIEDSVGNWYFFDQDGKMYENKHFVDVDSYGERG 1703
Db 1170 GSGFRRVGDDVOY--LSIGGYLAKNTFTQVGANQWYFDKNGNMVTEQVID---GKK- 1222
QY 1704 TYFFLKNQVSPRGL-----VQTDNGHY-----YEDNGKQVYRNQ 1738
Db 1223 -YFELDNGQLRHVLRGSGHGVYVDPKGVQAFNGFYDFAGRQDVRYFDGNGQMYRGL 1281
QY 1739 TINAGAMIYTLDENG 1753
Db 1282 HDMYGTTFYDEKGTG 1296

RESULT 11
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 15-Oct-1999
C:Accession: A44811; S22726; 528809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: A44811
```

A:Molecule type: DNA  
A:Residues: 1-1518 <GFP>  
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAA77900.1; PID:g47527  
A>Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
C:Genetics:  
C:Gene: gtfJ  
C:Superfamily: cpl repeat homology  
C:KeyWords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 22.2%; Score 2093; DB 2; Length 1518;  
Best Local Similarity 37.2%; Pred. No. 5.3e-91;  
Matches 511; Conservative 197; Mismatches 414; Indels 250; Gaps 42;

```
QY 557 TIKLTNDQYQALNGQLVLLRESKAADGSPSGDNTVTDQFSKNYATTCGDFYVKNVNGQ 616
Db 18 TIIVASVALATVJGGLSVT---TSSVSADETQKVTQNSNGTTLVTSPEATK----- 59
QY 617 VEPGSHATNQSKNDKDSOIIIVLNGKE---VKRQLVNDTK---EGAAAGFNDRDVKVNP 670
Db 70 -----EAKRNTKED---VLTPAKETNAVETATTTNTQATAEAATATTADV--AVA 118
QY 671 ATESSMSGFOGIITL-----PVTVKNEVQLVHRFSNDVKTGSGNVYDFWS 717
Db 119 AVPNK-----EAVTTDAVATEKAEBOPATVKAEV-----NTEVKAPEA----- 160
QY 718 ELPVKDSFKQNGPLKQFGLQTINGQQYIIDPTTQPRKNFLLQSGNNWYFSDSTGYG 777
Db 161 ---ALAKDSEVEAALSILK--NIKNDGKYIYVN--EDGSHKNEAITVNGQLLYEGKD-GAL 213
QY 778 TNALEQFAKGTVSSNEQYRNGNAAYSDYDKSTENYNGVLTADTYRKPQILKDGTTWD 837
Db 214 TSSTVTFPTGTIVVDFGSINNRAYDSSSEAFELDIGLTADSWTRPASIRDGVWQA 273
QY 838 SKETMRPILMWMPNLTQAYILNTMKHGNLLPSALPFFNADA-----DPAE-ILNYS 891
Db 274 STAEDFRLLMAWPNVDVQVNYLNTMSK-----VFNLDAKYSSTDQOETLKVA 323
QY 892 EIVQONIEKRISTGTNDWLRLMHDFVTNNPWNKDSENVAPSGTQ---FOGGFLAYENS 949
Db 324 KDQIKIEQIKQAEKSTQMLRETISAFVKTQPOWNKETENYKGGGEDHLOGGALLYND 383
QY 950 DLTPYANSYRLLGRMPIN-----IKDOT---YRQGBELLANDIDSNPNVVAEQ 996
Db 384 SRTPWANSYRLLNRATNTQGTIDKSIDLDESDPNHMGDFELANDVLSNPVVAEQ 443
QY 997 LNMVLYLLAFGTTTANNQANFDSVVDAPDNIDADLMIAQDYFNAAFGM--SDAVSNK 1055
Db 444 LNQIHYLMNWSIWMGDKDANFDGIRVDVNDVADMLQLYTYREYIYGVNKSANALA 503
QY 1056 HINILSDWHADEPYENKIGNOLTWDDTIKNSLNHGLS----- 1094
Db 504 HISVLEAWSLNDHYNKTDGGAALAMENKQRLALLSLAKPIKERTPAVSPLYNNTFTNT 563
QY 1095 -----DATNRWGLDAIVHQSLADRENNSTENVIPNYSFVRAHNNDSODQONAI 1143
Db 564 QRDEKTDWINKDGSAYNEDGVTKSTIGKYNEKYGD--ASGNYVETRAHDNNVQDIABI 622
QY 1144 I--RDVTGK--DYHTFTPEDEOKGDVADVQDQNSTVTKKNLYNPASVAILTKDITPRY 1201
Db 623 IKKEINPKSDGFTITDAEMKQAFIYNKDLSSDKKYLTLNNIPAAVAVMLQNEITTRY 682
QY 1202 YGDLTYDGGQYMEHQTRYYDTLTNLKSRVYVAGSQSMOT-----MSVGN 1248
Db 683 YGDLTYDGGQYMEHQTRYYDTLTNLKSRVYVAGSQSMOT-----MSVGN 1248
QY 1249 NNILTSVRYGKGAMTADTI--GTDETRTQIGVYVSNTPNKLGVNDKVVYLMHGAHKNQ 1307
Db 743 NEVITSVRYGKDIMTANDTEGSKYRTSGQVTLVANNPKNLNLDQSAKLNVEMGKTHANOK 802
QY 1308 YRAAVLTITTDGVINVTSDQAPVA-----MTDENGDLYLSSHNLYVNGKEADTAQGVAN 1363
Db 803 YRALIVGTADGINKFTSDADATAAGVYKVTGDSNGVLTFFGAND-----IKGYET 850
```

```
QY 1364 PDVSGYLAVWVPGASDQDARTAPSTE-KNSGNSAYRTMAAFDSNVIFPAFNFVYTP- 1421
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 851 FDMGFAVWVPGASDQDQIRVAPSTEAKKEGELILIKATEAYDSQLYEGFNFQITPD 910
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1422 -TKESRANVRIAGNADFFASLGFTSEMAPOYNSKDRFLDSTIDNGYAFTRDYDLGM 1480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 911 GSDPSVTNRKIAENVDLFKSWGVTSEMAPOFVSADGDTFLDSVQNGYAFADRYDLAM 970
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1481 SEPKNYGTDEDLRNAIQALHAKGLQVMDVPOQIYNLPKEVATVTRVDRGNVWKDAI 1540
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 971 SKNNKSGKEDRLDALKALHAKGIAQIADVWPQIYLPQKEVYVTRTRTDGAGRKIADAI 1030
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1541 INNLYVNTIGGG-EYOKKYGGAFLOKLOKLYPEITTKQVSTGVVAIDPSOKITWSAK 1599
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1031 IDHSLYVANSKSSGKDYQAYGGFEALAKYPEKMKVNMISTGKPIDDSVKLKQWKA 1090
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1600 YFNTNILHRSGGVVLKADG-QQYNIJGTTTKQ-FLPIQLTGEKK- 1642
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1091 YFNTNVLERGCVGLSDEATGKYF--TVTKGNFIPLOLTGKEKVIITGSDGKGITY 1147
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1643 -----GNGEGFVKGNDGNYVEYDLAGNMV----- 1674
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1148 FGTSGTAKSAFVTFN-GNTYIFDARGHMVTNSEYSPNGKDVYRFLPNGIMLSNAFVIDA 1206
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1675 VGNWYFF-----DQDG---KMVENKHFDV----- 1702
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1207 NGNTLYNSKGMQYKGGVTKFDVSETDKDGKESKVVFRVFTNBEVMAKGVTVIDGF-- 1263
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1703 GTYFELANGSVFRGLVQTDNGTYTFDNY-GKMYRNQITINAGAMIYTLDENG 1753
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1264 -TOYFGEDGFOAKDKLFTFKGKTYTFDAHTGNGIKDFTWRNGKWIYFDANG 1314
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 12

S22737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000

C:Accession: S22737; S28810; B44810; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 &lt;JAC&gt;

A:Cross-references: EMBL:Z21872; MID:947530; PIDN:CAA77898.1; PID:947531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377; PMID:1838351

A:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 &lt;GIF&gt;

A:Cross-references: EMBL:Z21873

C:Genetics:

C:Gene: gtfK

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:1456-1475/Domain: cpl repeat homology &lt;CPR&gt;

Query Match 21.3%; Score 2005.5; DB 2; Length 1599;

Best Local Similarity 42.6%; Pred. No. 8e-87;

Matches 449; Conservative 157; Mismatches 370; Indels 79; Gaps 25;

QY 742 NGQOYIDPTTGQPKNFLLQSGNNWYFDSDTGVTGNTALELQPAKGTVSSNGOYRNGNA 801

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 175 DGKYVLL-LEDGSHKKNFATVNGQVLYFD-ENGALSSTSTYSFTQETTLVDFDKNNA 232

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 802 AYSYDDKSIENGVNGLTADTWYRKQILKDGTTWTDSKETDMRPLMWPNLWPNLTQAYYL 861

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 233 AYDSTKASFELVDGYLRADSWYRKLEILACTTWKATKEDFRPLLSWHPDRDQTQAYL 292

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```
QY 862 NYMKOHGNLLPSALPFFFNADAPAEINHYSEIVQONIEKRISSETGNTDWTLMHDFVTN 921
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 293 NYMTKALSNGEETKDVFTIENSQASLSNAAQILQRIKIEVKIAANKSTDWLRQSIIEAFYKD 352
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 922 NPWNKDKSNVNFSGIQFOGGFLKYENSDLTPYANSYRLLGRMPIN-----IKDQYR 975
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 353 QDKWNINSPEKE--HFQKGALLFVNSDSTKWSNDYRKLNQATSYIKNHKIYVNSDG 410
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 976 GQEFLLANDINSNPVQAEQLNWLYLLNFQTTIT--ANNOQANFDSVRVADPNIDADL 1033
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 411 GYEFLLSNDIDNSNPVQAEMLNQLYIFFMWGOIYVFGDKDKDAHFDIRGIVADVNSVDM 470
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1034 MNAIQDYFAAAGM-DSDAVSNKHINILEDWNHADPEYFNKIGNPOLTMDDFIKNSLHNG 1092
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 471 LQIVSSYKAAKYVNESEARALANISILEAWSHNDPYVYNEHTAALSMDNGLRLSIYVG 530
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1093 LSDATNRWGLDAIVHOSLAD-----RENNSTENVVPIPNYSFVRAHDNNSQDOI 1140
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 531 LTRPVTNKGTTGA-RNASMKDILINGGVGLSNRAEVTYSVDLGFATYLFVRAHDS---EV 585
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1141 QNAIRDVTGK-----DYHTFTFEDEOKIDAYIODNSTVKKYNYLNPASVAILLTNK 1194
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 586 QTVIADIISKKIDPTDGTFTLDLQKQAFDIYNADMLKVDREYTHSNIPAAIYALMLQTM 645
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1195 DTIPRYVYGDLYTDGQGYMEHQTRYDTLTNLLKSRVYVAGQSQMOTMSVGG----- 1247
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 646 GAATRVYVYGDLYTDNGQYMAKSPYFDQITTLKARPKYVAGQTSYIHNLAGDGVSSAK 705
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1248 -NNNLTSVRYKGAMTATDT-GTDETRQOGIGVVVSNTPNLKLGNDKVVHLHMAAHKN 1305
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 706 DNKEVLVSRYGODLMSKTDTEGGKYGRNSGMLTIANNPDULKADGETITVNMGAHKN 765
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1306 QQYRAAVLTITDGVINYTSDGAP-VAMTDENGDLVLSHNLVNVNGKEADTAVQGYANP 1364
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 766 QAYRPLLGLTEKGIYSSLNSDSTKIVKYDAQG-----NLVFTADE-----IKGFTV 813
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1365 DVSQYLAVWVPGASDQDARTAPSTEK-NSGNSAYRTMAAFDSNVIFPAFNFVYTPK 1423
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 814 DMSGYLSWVVPVGTADDQNVLAKPSTKAYKEGDKYVSSAALEAQVIYEGFSNFODFYKE 873
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1424 ESERANVRIAGNADFFASLGFTSEMAPOYNSKDRFLDSTIDNGYAFTRDYDLGMSEP 1483
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 874 DSQYTNKLIANADLFKSWGITSFEIAPQYVSSKDGTFDLSIENGAYFTDRDYFAMSKN 933
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1484 NKYGTDDEDLNAIQALHAKGLQVMDVPOQIYNLPKEVATVTRVDRGNVWKDAIINN 1543
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 934 NKYGSKEDLRDALKALHKGIOVIADWVPDQYTLPGKEVVTATRTDTHGKVLDDTSLVN 993
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1544 NLYVNT-IGGGEYKQYGGAFLOKLOKLYPEITTKQVSTGVVAIDPSOKITEWSAKYFN 1602
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 994 KLYVTNKSNGNDFQAOYGGAFLOKLOKLYPEIFKEVMEASGKTIDPSVKIKQWEAKYFN 1053
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1603 GTNILHRSGGVVLKADGGQYVNL---GTTTKOFLPIQLTGEKKQNGEFGVKNDGNYFY 1659
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1054 GTNIQKRGSDYVL-SDGKLYFTVNDKGT---FLPAALTGTGDK-ARTGPAYDGTGYV-Y 1106
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1660 DLAGNMVKNVTFTEDSVGNWYFFDQDGKVENKHFVDVDSYGEKGYFFFLKNGVSGFRGLV 1719
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1107 TTSQOAKSQOFTYIN-GKQYFNDKGYLYVTGQITD-----GSNYFFFLPNGVMTDGYR 1159
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1720 QTDNGTYFEDNYGKMYRNQITINAGAMIYTLDENGK 1754
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1160 KNAKGQSLV--YKSGKGLATQGWKEVYKDDSGK 1192
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 13

B97033

uncharacterized protein, related to enterotoxins of other Clostridiales [imported]

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: B97033

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A: Reference number: A96900; MUID: 21359325; PMID: 21359325

A: Accession: B97033

A: Status: Preliminary

A: Molecule type: DNA

A: Residues: 1-2817 <KUR>

A: Cross-references: GB: AE001437; PID: AAK79053.1; PID: g15023993; GSPDB: GN00168

A: Experimental source: Clostridium acetobutylicum ATCC824

A: Genetics:

A: Gene: CAC1079

Query Match 4.58; Score 424; DB 2; Length 2817;

Best Local Similarity 19.78; Pred. No. 1.2e-11;

Matches 452; Conservative 277; Mismatches 766; Indels 796; Gaps 125;

36 AHADQVQAASTTQDSTVNDNDKTVLALD---TNTDQSAQTDDKKQVVSNTNOSKTDD 92

183 SNTDIKTTEAPANAFIKDTPNNQSDSALAKNALSNNLAADSSQTSKYTSNNDAKPVN 242

93 TSTADKNSTSPVSVLFSNTE-----KQAKNY-----NEQDKNGY-NIDTAYFSNNQLH 142

243 TTSDDKASNL-----NNDSDQGWTKDGGKYYVNGVQOQFQSINKSIYYFNDGSM 296

143 VSGWNAATNAS-----QGTNSRQIIVRDITTNELGRITDVTNNVAREPDVKNV 188

297 QIGWLKYNNSYFSDASGVMLTGLQNINGTYGEND--DGKLLTGLQAINN----- 345

189 HNVYNADNSGPDVNVNIDFSKMKDYRDS-----TEIVSRYSGNKK----- 228

346 -NYIFNNDGVMQGTWITCNDISKYFDNNGVMQGLVHNKKYGFNGDGLLGLQAIN 404

229 -----SVDWMSQPTIFOKNNVAYLDTPEVKNELHATG--WNAFN----- 266

405 NYTYFDSNGVMQTDW-----ITIDGSKYFVSVNGVMQGTIIYISGYGFANDCKLLTGL 460

267 SAINVNHHEVILFDOTNG-KEVAQOEVBEGQ-----SRPDV-----AKVPQVYVGA 312

461 QVINGNSYI---FD-TNGIRLVSWITIDGKDYFQNDGILTDNNVYDGRKYFYISGVK 516

313 NSGENVTNIDLDYTHOYVLSRYSNSDNGEGNVTVWENPQSJAPAN-----OSNOGYL 368

517 QTG-----LONIDGNYIYFSSGIMQTLGKIDKTIYFGDNGIRQIGWTYQNNKIYF 570

369 DSFDISKNGEVTVTGWNA-----TDLSELQNNHYVILFDOTAGKQV 409

571 NS-DGSMQTDLEKISYSFSPYHYQYFGPDGDKLLTGLQTIKGNFY--YFDSNGISQM 627

410 ASAKADILSRPDVAKAYFTVKTATNSGFKVTFKYNLL-OPGHQYSVYSRFSADENGND 468

628 GWNIDGKDFYNSIMTENWVIND-E-KYFYINNKKQTGFQYI-----NGK- 674

469 KRHTDYFSPVILNOTASNI---DTITMTSNGLHAGMA-----SDNSI----- 510

675 ----YIYFDPGIMQGTQITISGNTIYLLDNGVKQTGWVTKIKDYFDGNGVMINYWF 730

511 -NETTPYAILLNKEVTRQKMSLTARPDVAAYVPSLNSAVSGFDTTIKLTNDQYQALN 569

731 DNDKTYIYI---NGNMQP-----GAI---SINHHYGFDDN-GIMQGTGWRIN 771

570 GLOQLVLLRFSKAADGNPSGDNVTIDQFSKNVA--TTG---GNEDYVKVNGNOVEFSGWA 624

772 GRITYFDNGAAGKGLVYEGK-TYFYNTYAYLDTGFIFNNNYIFLDNNGVVRTGW-- 828

625 TNQSNK---DSOWIIV---LVNGKE-----VKRQLVNDPKEG----- 656

829 INYSNNRYLDSGTGVRVTGFTIDGNKYIFDSSGAMCTSFITVNGNTYGFSGDKGIMLTGW 888

657 ----AAGNRNDVYKVPAINSSMSGF-----OQIITLPVTKNENVLVH 699

889 QTLSSNYSYNIYFNS--DGSQAKGFFTYLKGTYFEPNPGYMLLGYNYNGK---YY 943

700 RFSND--VKTG-----EGNYVDFWSELMP---VKDSFQKNGPLKQF----- 736

944 YPDNDGVIQTGWTDTRSSKYILD-----PSGAATGFGQNINGDKYFYFNSSGIMQGLVY 997

737 -----GLOTINGQOYYIDPTIQCPKNEL-----LQS 763

998 VNPDIYGFDDNGHILTMHSINGIYIFD-STGAQKAGFVYLGKTYFYFNTNMYTGFWNA 1056

764 GNNWIFDSD---TG-VGTNALELOF-ARGTVSSNEQYRNGNAAAYSYDDKSIENYNGYL 817

1057 NNNLYFDNEGVMQGTWINNSNRYVFSATGASVTGFQIDGN-KYCFDS-----NGAI 1109

818 TADTWIRPKQILKDGITW---TDSKETDRPILMWVWPNTLTQAYLYNLMKQHGNTLLPSA 874

1110 YTDV-----VTINGSTYGFNTDG-----IMLTGMQIRYNRGYSSYFNTY----- 1149

875 LPFENAD-----ADFAELNHYSEIVQONIEKRISSETGNTDLWLT-- 913

1150 --YFNSDGTAKTGFTYLNKTYFNPSDGRMLQGYQYINGNHVYFAPDGTMTGTWITINGS 1207

914 -----LMHD--FVTNNPMWKNKDSENVAFSGIFOG 941

1208 SKYYLDPGAAVTGLQTINGNKYCFDSNGILOHNGIFYIGNTYYGSDNNGIMLTGLQLIN 1267

942 GFLKYENSND-----LTPYAN-----SDYRLGLRMPINIKDQTYRGQEEFLANDIDNSNP 990

1268 GYLYCFNSDGSVKTGLVYLGKTYFYFDSYVSQFQ--NINNTY-----YFGND-----G 1315

991 VVQAEQLNLYLLNFGTITANNDOANFDSVRVADPNIDAD-----LMNIAQ 1038

1316 TMTQGWNYGY-----RYLNDSGIKVTQWTIDGNKYFYDYGAKTGIYNDIG 1365

1039 DYFNAAYGMDSDAVSNKHINILEDNHADPE-----YFNKIGN 1076

1366 NY---YGFNNSGV-----MLTGWHINGSTYFENSNGIANTGFTYLGKTYFDSYGR 1415

1077 PQ---LTMDDT-----TKNSLNHGLSDATNR-----WGLDAIVHQ---SLADRENNSTENVV 1122

1416 MQGSMTINGSTYFYFANGVMKTSDBSPNTLAVGWVRDYSYIYYIYNAAGTKLTGLQTI 1475

1123 IPNYSFVRHNDNSQDQIONAIRDVTGDKYHTTFEDEQKIDAY-IQDNSTVKYKNLY 1181

1476 GNTYF-----DSNGIMQGTIITINGNRYGFGV-----NGVMYLGLOFINNTYYSNSY 1524

1182 NIPASVAILTNKDTIPRVYVGLYTD----- 1208

1525 GISQTFVTLG-----NTYFDSYGEMLGLAYINNYYFNSKGMETGWSYLRAN 1579

1209 -----GGQYMEHQTPRYDITLNLKSRKYVAGG-----QSMQTSVGVGN 1250

1580 PNGILLTGFTINGKTYIFNSDGLLYD-LQYINGSYGFDKNGVMYLGLOQT--IGNTY 1636

1251 ILTSVRYGK-GAMTATDTGDETRTQIGVWVSNTPNLKLGVNDKVVLHMG-AAHKNQOY 1308

1637 YLNSNGISQSGFITLNGKTYFDSYGMRTGIONINNYYFEGDNGTLOTGWISQDNLYR 1696

1309 RA-----AVLTDTDG-----VINYTSQCAPVAMT-----DENGDLVSSHNLYVNGK 1351

1697 YANSSGYCLTGLQTIIDGKYYFNSYARMETGLYINNTYFGDNDGTLLYSWHN--INGR 1754

1352 ----EADTAVQGYAN-----PDVSGYLAIV-----WYVPVGA 1378

1755 MYCFNDGTVKTGWINYLGRCVYLDSSQGSFLTGLLTIGHNIYYFSDYSMTGTWTS 1814

1379 SD---NODARTAPSTKNSGNSAYRNAAFDSNVIFEAFSNFYVPTKESERANV----- 1430

1815 SKIYFNEGIMLGFTQIDGNTYI-----FDSYGN--STGTASINGNCYGFND 1861

1431 -----RIQANADF-----ASLGFTSFEMAPQYNS--KDRTFELSDTNGVAFD 1474

1862 DGIMLGQWTISGNNYFNDPGTAKIGLNTYEGKTYFSTGGVYTGQGIININSNTYF-- 1919

1475 RYDLGNSE---PNKYGTDEDLRNAIQALHAKAGLQVMADWVPDQIYNLPGKEVATVTRVDD 1531



Db 1920 GYDALKTGWRNYYIYADNNGIIQ-----TGKLTIDG--KNVLFNYSG----- 1962  
 Qy 1532 RGNWVKDAI--INN-----LYVNTIGGYSQYKYGGA-----FLDKIKLILPEI 1575  
 Db 1963 --IMNGIQSINNYYGFDNGMLFGVHSINGQYTYFNADGSVKGTWIPYEGKMY-- 2016  
 Qy 1576 FTRKQVSTGVAIDPSQK-----ITEMSAKIF-NGNINLHRGSGVYLKADGGQ 1621  
 Db 2017 YANPSYCTGATINNNTYYFDNNGAMETGIITVDNKKYIINSYGI--RETGYKI-IDGKT 2073  
 Qy 1622 YYNIGTTTKOFLPIQLTGEKKQGNFVKNGDNGNYFYFDIAGNWKNTFIEDSVGNWYFF 1681  
 Db 2074 YY-----FNPNYSGIMQTG-----VISINGNYGFD--DNGVMQIGLQKLNNTYFF 2118  
 Qy 1682 DQDSKMW-----ENKHFVDVDSY--EKGTT-----YFEELKNGVSPRGGJLVQDNG- 1724  
 Db 2119 NSNGTATGTTTGLSNKYYPNPSYGAEEVIESINGHYIYFNNGIMQTG--VORINGY 2176  
 Qy 1725 -----TYFYDNYGKMVRN-QTINA-----GAM----- 1745  
 Db 2177 VICFNNDGTAAFGQSINGNTYVLYNYGCTYGTGWTINSNRYFNSDGVMLTGAQNTIGY 2236  
 Qy 1746 IYTLIDENGLI 1756  
 Db 2237 IYGFDSNGIML 2247  
 RESULT 14  
 S32920  
 cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 21-Jul-2000  
 C:Accession: S32920; E69730; T47101  
 R:Foster, S.J.  
 Mol. Microbiol. 8, 299-310, 1993  
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis  
 protein.  
 A:Reference number: S32919; MUID:93302506; PMID:8316082  
 A:Accession: S32920  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <POS>  
 A:Cross-references: GB:L05634; NID:g304177; PIDN:AAA22893.1; PID:g304179  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauele  
 y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, K.; Uchiyana, K  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: E69730  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <KUN>  
 A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CABL15959.1; PID:g2636469  
 A:Experimental source: strain 168  
 R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
 Microbiology 141, 337-343, 1995  
 A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai  
 A:Reference number: Z24350; MUID:95219088; PMID:7704263  
 A:Accession: T47101  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <VOS>

A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782  
 A:Experimental source: strain BGSCIA1  
 C:Genetics:  
 A:Gene: wapa; N17G  
 C:Superfamily: Cell wall-associated protein wapa

Query Match 3.7%; Score 344.5; DB 2; Length 2334;  
 Best Local Similarity 19.0%; Pred. No. 5.2e-08;  
 Matches 458; Conservative 271; Mismatches 760; Indels 917; Gaps 121;

Qy 4 KKHFLKXSGKQWTVAAVATVAVSTALLYGVGAHAAQQVQAS--TTQDQSTVND--TD 60  
 Db 6 RNFRRF-----IAAFLVLAISMISLPADVLAKESTEENGRIAAADPPETLQKEOTE 58  
 Qy 61 KTVALDTN-----TDQSAQTTD-----KKQV-----VSNTNQSKTDDTSTAD 97  
 Db 59 EAVPDPKIDNKEGILTERTENTKLYYEGDGVYQEVYLDPIHTKETPDADWEDISP 118  
 Qy 98 KNSTSTPV---SVLPNNTEKOAKNYNEODKNGYGNIDTAYFSNNQLHVS-----GWN 147  
 Db 119 KESTSKQVETENAILNSDFQKMKN-----GLY-----ATFEHNDHRTVYSLAEAKGP 167  
 Qy 148 A-----TNASOGTNSRQIIVRDITTTNNELGRDVTNNVARPDVKNVHNYNA----- 194  
 Db 168 KTSLPKQTSADYKTDSEIYVDPVFNIDLTQTFNENIKEDLVLHQYNGYNTFTFOLK 227  
 Qy 195 -----DMSGF 199  
 Db 228 TDLQAKEQEDSGIDFSDKGVKVPKPFMTDSKLDSELSGEVRSKVSYKLEKNEGY 287  
 Qy 200 DYNVNIIDSKMKD---YRDSIE-----IVRSY---SGNGKSVDMWSQDIT-- 238  
 Db 288 LLHLADENLWLDKPPRVPSIDPSTLSVSSDTPFWSAYPTTNSASSQKDWANLKAYV 347  
 Qy 239 -----FDK---NNYAYLDTFEVKNGE-----LHATG-W-NATNSAI 269  
 Db 348 LKTGYDKTTGTNYAFMKFNFKLQNNMTVTKATLTKTYVAHSYVTKATGLWLDTPVNS-- 405  
 Qy 270 NYNHEFVILFDOTNGKEVARQVREGQ---SRPDVAKVYQVYVGAANSFNFTNISDLDY 327  
 Db 405 NYDNKAVTWTNTPASKNGTKGADVHKQWASVDVTAAVKSWNSGGANYGFKLHTNGKREY 465  
 Qy 328 THQYQVLSYNSDNGEGDNVTYFNPSIAPA-----NQSNGQYLDSEFDS----- 374  
 Db 466 ---WKKLISSANSANKPIEVTYTI--PKGNTPTIKAYHNGDSTGY---FDISWKKVEGAK 518  
 Qy 375 -----KNCV-----TYTGNW-----ATDISE 391  
 Db 519 GYKVIYNGKRYQATISAGNVTSWSTGKKKIWPTSAFIASKRYKHLHDGKGAELADPS 578  
 Qy 392 LQNN-----HYVI-----LFDQTAGKQVASAKADLISRPDVAKAY-PTVKATNSG- 436  
 Db 579 YKNSGGSVATSKNYWIGVSAIFDQEGAMSAPAKPVI---PNVGKAQAPSAGYNNNGNA 635  
 Qy 437 ---FVYTKVNNLQPHQYVSYSRFSADENGNGDKRHTDYFSPVIL-NQTASNIDIT 492  
 Db 636 TCYFOLSMKAVSGATGYKQVQV-----NGKG-----FETLDIGNOT-----S 672  
 Qy 493 MTSNG-----LHIAGWMASSNSINETTPY-----AIIIN 521  
 Db 673 WTKGKKIWPSTSAEIKAGKYALHLKDGSCAELPINPGPTYKNAGGAKRNVSPKIIAYN 732  
 Qy 522 NGKEVTRQKMSIATRPDVA--VTPSLYNSAVSGFDTIKLTNDQYQALNGQLVLRP 578  
 Db 733 KQGEIATASPAATPALPDIAEPKNVNTGYLYTNTKSSQTGYVNLINKVQNAKG-YKVNLYN 791  
 Qy 579 SKAADGNSGDNVTVDQSKNVATY-----CGNFD-----YKVNQNG 616  
 Db 792 GREYQSFVGDADHWTTQKNKNTWPTSEBKAGSYKLHDTGKGGLALDPSVYNNANGNY 851  
 Qy 617 VEFSGWHATNQSNDKSDQWIIIVLVNGKRYKROLVNDTKEGAAGFNRDVYKVNPAIENSS 676  
 Db 852 GKGNVSYFTLVAYD-----ANGETIPTAPFNTFHEGAEEFLGTEY----- 892



```
QY 677 MSGFQGIITLPVTVKNEVQLVRFNSNDVKTGEGNYVDFHSELMPKVDKSGKNGPLKQF 736
Db 893 -----WSIIDIP-----SQGLNGATGCVI-----VNEEDLSIDGRGP--GL 926
QY 737 GL-QTIN-----GQQYVIDTTQPRKRNFLQSGNNWIYFDS-----773
Db 927 GLSRTYNSLSSDHLFGGQVADAETS-----VISTDGGAMYIDEDATTHRTKKADGT 960
QY 774 -----TGVTNALEL--QPAKGTVSSNEQY-----RNGNAAYSDDKS-----809
Db 981 YQPTGYVLELTETADQFILKTKDQTNAYFNKGGKLOKVVDGHHNATVYVNDKNOLTA 1040
QY 810 -----IENVGYLTADTWKPKOI-----LK-----DGTWTDKSETDMRP 845
Db 1041 ITDASGRKLTFTYDENGHVTSITGPKKKKVTYSYENDLLKVVTDGTVSYDYDSGRL 1100
QY 846 ILMVWPNWTLQAVYLNMKOHGNLPSALP-----FFNADAPPAELNHYSEIVQON--897
Db 1101 VKQYSANSTEAKPVTEY-QYSGHRLKKAINKAKETVYVSYDADKCTL-----LMTQPNGR 1155
QY 898 -IEKRISGTG-----TWLRL-TLMHDFVNNPMKNSENV-----N 933
Db 1156 KVQGYNEAGNPIQVIDDAEGLKITNTKYEGNVVVEDVPNDVGTGKATESQYDKDGN 1215
QY 934 FSGIQFGGFLKYE-----NSDLTPYAN-----SDYRLLRMPINIKDOTYR-----975
Db 1216 VTSVKDAYGTETEVYKNNDVTMKMDTEGNVTDIAYDGLDAVSETDQSGKSSSAAYDYK 1275
QY 976 QGEFLANDIDNSPVV-----QAEQLNWLYLLNFGTITANNDQANFDSV-----1021
Db 1276 GNQIOSSKDSASTNLLKDGFEACKSGW-----NLTASKDRRKISVIADKSGVLSG 1327
QY 1022 -----RVAPDN-----IDADLMNTAQDYFNAAVGMDS 1049
Db 1328 SKALEVLSQTSAGTDHGYSSATQTVLEPNTYVTLSSKIKYDLAK-SRAYFNIDL-RDK 1385
QY 1050 DA-----VSNKHIL--EDW-----NHADPE-----YFNKIG 1075
Db 1386 DOKRIQWIHNEYSALAGKNDWTKRQITFTTPANAGKAVVYMEVDHKDKGKAWFDEQV 1445
QY 1076 NPOLTMDDTIKNSLNHLGSDATNKGWGLDAIVHQSADRENNSENVIPNYSFVRAHDNN 1135
Db 1446 LEKEVSSSYNPVQNSFTSATEWNN-----VSGASVSEEGFNDDV-----SLKAATSA 1496
QY 1136 SQ-----DOIQN-----AIRDVTGKDYH-----TFTFDEQKGIDA 1166
Db 1497 SQAGSVTKQTWVLGOSANDKPVYLLTGMASKASSVKFTDEKDYSLQANVTYADGSTGT--1554
QY 1167 YIQDQNSTVKYN--LYNIPASYAI-----LLTNKDIPIRVYVYGLYTDGQYMEHQTR 1218
Db 1555 YNAKPPSGTQBNRAAVVPIKTKPINKVDISILFQKSGATGVTWFDLRLIEGSLITKST-1613
QY 1219 YYDILTNLKSRVKYVAGGQSMQMSVGGNNILTSVRYGKAMTATD---TG--TDETR 1273
Db 1614 -YDS-----NGNYVTKEDEELGYATSDYDETKKTSSTD 1647
QY 1274 TQIGVYVNSVTPNLKLVNDKVVLMHGAAHKNQYRAAVLTITDGVINYTSQCAPYA-M 1332
Db 1648 AKG-----EKT-----TYTYDQADQUTNM 1666
QY 1333 TDENGDLVSLSHNLVYNGKEEADT-----AVQGYA-NPDVSGYLAVWY-PVG---ASDNQD 1383
Db 1667 TLSNGTSLHSYD--KEGNEYSKTRIRAGADQTYKFEYDVMGKLVKTTDPLGNVLASEYDA 1724
QY 1384 ARTAPSTEKNSGNSAYRTNAAFD--SNVIFAFSNFVYTPTKESERANVRILAQNADEFAS 1441
Db 1725 NSNLTKTISPNGNEVSLVSQDTRVKSYSNGTEKIFYTDYDKNGNETSV-----1773
QY 1442 LGFTSFEMAPQYNSKSDRTFLDSTIDNGYAFTDRYDLGMSBPNKYGTDEDLRNAIQALH-1500
Db 1774 -----VNKEQNTTKKTF-----DNKNRLTELTDRCGSGTWTYTPSDSKLKTFSWIHG 1821
```

```
QY 1501 -KAGLOMADWVPDQIYNLPGKEVATVTRVDRGNWVK-----1537
Db 1822 DOKGTNQFTYNLQDQMIEMKDDSTSSYFYDENGNVQVQFITNGGGTSEFSYDERNLVSSL 1881
QY 1538 -----DAIINNLYVNTIGGVEYQKYGGAFDLKLOKL-----1571
Db 1882 HIGDKNGGDILTSEYBDA--NGNRTTINSSASGKVQVEYG-----KLNQLVKETHEDGT 1934
QY 1572 -----YPEIFTKKQVSTGVAIDPSQKITWSAKYFNGT-NILHR-----GS 1611
Db 1935 VIEYTYDGGGNKRTVTT-----IKGSSKTVNASFINMQLTKVNDESISYDKNGN 1985
QY 1612 GYVLKADGGQY-----NLGTTTQKFLQTLTGKKGKOGNEGFVKGNDGNY--FYFDLAG 1663
Db 1986 ---RTSDGKFTYDWAEDNLATVK-----KGEDKPPFATYKYDEKG 2023
QY 1664 NMVKNTFIEDSVGNWYFFDQDGKMWENKHFVDV-----SYGEKGTFFFLKNGVSFPG 1716
Db 2024 NRIQKT-VNGKVTNYFY---DGSLLNVLYETDADNNVTKSYTYGDSGQL-----2068
QY 1717 GLVOTDNGTYFEDNYGKVMRNQITNA-GAMIVTLDENGKLIRASYNDA-EYPTSTDVGK 1774
Db 2069 -LSYTEGKKKYFHY-----NAHGDIILAISDSTGKTV-ACYQYDANGNPTKTEASD 2117
QY 1775 MLDQNK 1780
Db 2118 EVKDNR 2123

RESULT 15
F82865
Hypothetical protein UU482 [Imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82865
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: the complete sequence of Ureaplasma urealyticum: Alternate views of a
A:Reference number: A82870
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AE002145; GB:AF222894; NID:g6899476; PIDN:AAF30894.1; GSPDB:GN
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU482
A:Genetic code: SGC3

Query Match 3.4%; Score 321.5; Db 2; Length 4688;
Best Local Similarity 18.8%; Pred. No. 1.9e-06;
Matches 410; Conservative 328; Mismatches 762; Indels 681; Gaps 118;

QY 39 DQVQQAQTQDQTSVNNNDTKYVALDNTDOSAGTTDKKQVSVNTQS-----KTDD 92
Db 1369 NQTLHLKDLTLTHQFSI--NPSNNAVSLKNNNTIEI---TKRILVNNDOQLSAKIEVD 1422
QY 93 ---TSTADK-----NSTSTPVSVLPSNTEKQAKNYN---EODKNGYNIDTAYFS 137
Db 1423 IDNVLTADKPNVIOLENSD-----ANNLKLATNAQVVENNK-KFLKFDLVSLK 1474
QY 138 NNQLHSGMNAATNAGTNSRQIIVRDITN-----NELGRDVTNNVARPDKVNVHNY 192
Db 1475 INQNYVIKEISFN-SKPTNAYFNFTNNKTNIVSYDEQNKISLSNNINPTSYK--PQT 1531
QY 193 NADNSGFDVNVNIDFSKMDYRDSIEIVSRYSNGSGSVDMWSQPIFDKN-----242
Db 1532 NKXDKSVNIDVDIQVDKQLLANQYLRLKQLNDNKTIV--WTDPIILFNNAKISFKLSNL 1589
QY 243 --NVAY----LDTFEVKNGELHAT---GWNA-----TNSAI-----269
Db 1590 IHNRAVELEGLYYFDQNSVNDNTNNQISFNKHKPIEFESLTIYDVTNNAIKTVS 1649
```

QY 270 --NHNHVFILFDTNGKEVAROEVRGOSRP-----DVAKVYPQVVGAAANSGFN---VTF 320  
Db 1650 AHNAQVHFKL---KTNDLEANDQIIEAVFAPTNNDLNDQKVYEAALNNVTNSFNEGELEF 1706  
QY 321 NISDLTYHOYQVLSRYNSDNGEGDNYVWFNPQSIAPANQSNQGYLDSFD:SKNGEV- 379  
Db 1707 NLSGLRETTYRLI-----KVTF-----KKNPKAY---ELLNKGNVIF 1742  
QY 380 -TVTGNATDLSLONNHVILFDQTAGQVAKAKADLISRPD-VAKAY----- 426  
Db 1743 EYKNGSOAYEFTQKPEKVI---DVVSSSTNTTQOEIIVKIDIGORANNKKLELVYES 1800  
QY 427 -----PVTKTA-----TNSGFKVTFK-----VNNLQPHQYSYVSRSFADENGNDKRRH 471  
Db 1801 NILGDPBEIKTVDNNNSVHLSFDKKEYNLVLNKLPGRRYS-LKKINIKVEVDNGQDHE- 1858  
QY 472 TDWFSVILNQ-----TASNIDTITMTSNGLIHAGWASDNSTNETTPYAILN 521  
Db 1859 ---FVKEINVNFDFVNLQSEITASSVEEN-----DRAFDKLNQTT-IKINLK 1903  
QY 522 NGKEVTRQKMSLTARPDAVAPYPSLNSAVSGEDTTIKLINDQYQALNGOLQVLLRFSKA 581  
Db 1904 DDNDILKTN-----DIATI---TYDNE-QKVDAIVK-TNAQNKYLEAITNLVFNK- 1950  
QY 582 ADGNFSGDNTVTDQFS-KNYATTGGFDYVKVNGNQVEPSGWHATNQSNDKDSQWILVLV 640  
Db 1951 -----NVIKIEFKLSQT-----FKVGKNN-----TNVIYDESNNKLIIN 1989  
QY 641 NGKEVKRQL---VNDTKEGAAGFNNDV-----YKVNPAIENS-----SMG---FQ 681  
Db 1990 NDFQIIGIPLSTDVNSQNIIVAN-NKHEVISTTDLDFKVNPHISKLKFKLPQNINGEVWYS 2048  
QY 682 GIITLPVTNKNVQLVHRSNDVKTG-----EGNVVDVFWSELMVPKSPQKNGPLQK 735  
Db 2049 PILNLSIIYNNKNVNFILDLNKLNLQVRLVDVYIIDNNNTINDNNKVPKANVTRI 2108  
QY 736 FGLQTINGQOYYIDPTTGQPRKNFLQSGNNMIYFSDTGVTGNALELOFAKTGVSNSQ 795  
Db 2109 IDI-----APGKTTISKNTW-----NTTSSQOFEFVINSDDGNEV 2146  
QY 796 YRGNRAYSYDD-----KSIENVGILFADTWYRPQKILKGTWTWDSKEFDMRP 845  
Db 2147 LDNLEATISFKGOTLTPVKVNIQKNNYL-----IKGOITNLEPENRYVLE 2195  
QY 846 ILAVWVPNTLTQAYLYNMYKQHNLPALSPFPFNAD--ADPAELNYSIEVQONIEKRIS 903  
Db 2196 SILLAKPNKTK-----PLVVEILNKDDISFQTQAGNYKVI---QIKSQNP 2238  
QY 904 ETGNTDMLRLMHDFTVNNPMWKNKDSENVNFG-----IQFGGFLKYE----- 947  
Db 2239 STVDTKORIKLKDGIQN-AWNEKQLEITYSANDNSTKTAIIKLEKNKLEVEFELTNLE 2296  
QY 948 -----NSDLTPYANS-----YRLIGRMPI---NIXDQYRGQEFLLA-- 982  
Db 2297 KNRITYFTKIELINDNNKTFFPKNSIQDKFIVLSNNGVGVGNIIEIQDRVNHLSAK 2356  
QY 983 -----NDID-----NSNPVYQAEQL---NWLYYLLNFGTITANND-----QAN 1017  
Db 2357 IRFELNDLVNSDEQATITYNNTQTSKAVITDQNKYLEATFSLNLVNLNKTIINKTE 2416  
QY 1018 FDSVRVDAPDNIDADLNLNIAQDYFNAAAGMD-----SDAVSNKHINI-----LED 1062  
Db 2417 FNTKPRNASKNIGINDTNVZYDATNLIINNDLKITGLPHTLKEFEANNTN:SVSLELDT 2476  
QY 1063 WHHADPE-YF-----NKIGNP-----QLTMDDTIKNSLNHGLSDAT-NRW---G 1101  
Db 2477 NNHLSKNLYFIAPKPSDNGGSVLTNFIADKIVTNNNNKKELTFLNLTNSRQYIFKG 2536  
QY 1102 LDAT--VHQSLADRENNSTEN-----VVIPNYSFVRAHDN-----NSQ 1137  
Db 2537 LYVYNSNOTNIDENNNKFEKNSVDYKITYKPTITTIQKNGNWTFFQPNQAQKFEKINSD 2596  
QY 1138 DQIQ---NAIRDVTKGDYH-----TFTFEDQKGIDAVIQDNSTVKKNLY----- 1181

Search completed: April 23, 2003, 15:10:53  
Job time : 97 secs

Db 2597 ENVFESTOLDATIIPSDOHDIKKITTITKLKQKDNQWYIEDISLAYNDTTLKDDISITK 2656  
QY 1182 --NIPASYAILLTNKD-----TIPRYYG--D 1204  
Db 2657 PMNAFANLKIQIENKEQISFTTQSGPTQLVSTTSHDLMWNDQPNASNOQTITAKVSGVND 2716  
QY 1205 LYTDGGQW--EHOIRYYDTLT-----NLLKSRVKY-----VAGGO-SMOTKSVGNN 1249  
Db 2717 LYNNKKILVYRNNOKSLVESNELTLQDQOYITFLPISVANRQYSPFEKIKITSN 2776  
QY 1250 NILTSVRYKGAAMTADTGTDETRTQGGVYVVSNTPNILKGV-NDKVYVHMGAAHKNQY 1308  
Db 2777 NNFETLNNSTNTVNSFNVNASKTQ-----IVIDNT-NLATNITSTSATITKLSKDHVF 2830  
QY 1309 RAAVLTITTDGVN---TSDQAPVAM-----DENGDLVLSHNLVVKKEADTAV- 1358  
Db 2831 QVG-----DSIITYLKSNDDEKQEISYTKTITSISSDSEATVSNFNTDQTLKEANYKLI 2895  
QY 1359 -QCYANPDVSGYLAVVVPVGSADNODARTAPSTKNSGNSAYRTNAAFDNVIFEAFSNF 1417  
Db 2886 KVGFPVKPTLAYTNI-----NNDANNVIFEDNNSN---YNFKTLIVDHKVTNVSSND 2934  
QY 1418 VYTPKTESERANVRLAQNADFPASLGFTSPFMAPOYNSKORTFLDSI-----DNGYAF 1472  
Db 2935 SINTTQVNIIDIGIQT-----WINKIQLVITSNDGEEILSDOKTLLWANNHYSF 2987  
QY 1473 -----TDRYDLG---MSEPNKYGTDEDLRNAIQALHKGALQVMAOW----- 1510  
Db 2988 ELSNLKHNKRYTLKEVRIINDDNKTSIIFHLKNGI-----ADWFTVNTKQTSISI 3036  
QY 1511 -----VPDQIYNLPGKEVATVTRVDDRGNYW---KDALIN--NNLYV---VNTIGGGEY 1556  
Db 3037 SSIIEPIARAKNNLOSTQIRFI--LNDPDNVLNSNEEAIINYGENLSVKAKVKVIGSOKY 3094  
QY 1557 -OKKYGGAFPLDKLQLYPEIETKQV--STGVAIDPSQKITESAKYFNGNINILHRSGY 1613  
Db 3095 LETTFNNLALNODTIINSITTFINKPIKAATNIGIDNSNKI-----YNNNTNIN- 3142  
QY 1614 VLKADGGQYNNLGTTKQPLPIQLTGEKK---QGNEGFVKGNDGNYFYDLAGNWKNT 1669  
Db 3143 PLKID-NNFVSIG-----PVQSDTNKRITLEYNNKTNVKVN-----LEKAN----- 3183  
QY 1670 FTEDSVGWNVFP-----DQDGKMWENKHFVDVDSYGEKGTFFLKNGVSRFGLVQTDNGT 1725  
Db 3184 --DNILQNLIFYATFKDQNNETVKSS-IVSYSSIKKEGK---KQIIELRIPNLKS-NRL 3235  
QY 1726 YFIDNY-----GAMVR-----NOTINAGAMITYTLDENKGLIKA 1758  
Db 3236 YHFDNLYLLNQDDKDLIDNNKLIIRSNVNVABIEVPGNTTFKKNNVEQINSNSVNFKL 3295  
QY 1759 SYNDAEYPTSTDVGKMLDON 1779  
Db 3296 QVESDDE-----DILDNN 3308

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: April 23, 2003, 15:01:06 ; Search time 32 Seconds  
(without alignments)  
2308.417 Million cell updates/sec

Title: US-09-995-749A-2

Perfect score: 9424

Sequence: 1 MEIKKFKLYKSGKWTA.....SDAEYPTSDVGRMLDQNK 1781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Query Score	Match	Length	DB ID	Description
1	2446	26.0	1462	1	GFCD_STRMU
2	2347.5	24.9	1375	1	GFCD_STRMU
3	2330.5	24.7	1476	1	GFCD_STRMU
4	2274	24.1	1592	1	GFCD_STRMU
5	2271	24.1	1597	1	GFCD_STRMU
6	2191.5	23.3	1365	1	GFCD_STRMU
7	344.5	3.7	2334	1	WAPA_BACSU
8	287.5	3.1	1861	1	APU_THETU
9	286.5	3.0	1296	1	ASAL_ENTFA
10	285.5	3.0	2003	1	YDBA_ECOLI
11	276	2.9	2660	1	YBEJ_ECO57
12	274.5	2.9	1656	1	OMP_B_RICJA
13	269	2.9	1655	1	OMP_B_RICCN
14	266	2.8	1933	1	BGA_SALTY
15	264.5	2.8	1645	1	OMP_B_RICTY
16	262	2.8	1902	1	PJP_LACLC
17	261.5	2.8	1654	1	OMP_B_RICRI
18	257.5	2.7	2358	1	YBEJ_ECOLI
19	256.5	2.7	2710	1	TOXA_CLODI
20	255	2.7	1902	1	PJP_LACLC
21	254.5	2.7	1902	1	PJP_LACLC
22	248	2.6	1577	1	HLVA_PROMI
23	248	2.6	2366	1	TOXB_CLODI
24	246	2.6	1643	1	OMP_B_RICPR
25	246	2.6	1902	1	PJP_LACPA
26	245	2.6	1256	1	ATL_STAAU
27	239	2.5	1279	1	APU_THESA
28	233.5	2.5	1423	1	FRUA_STRMU
29	233.5	2.5	1532	1	IGA_NEIGO
30	232	2.5	2249	1	OMP_B_RICRI
31	229	2.4	1018	1	FNBA_STRAU
32	225	2.4	1770	1	YCB9_YEAST
33	224.5	2.4	1545	1	IGA3_HAEIN

34 221.5 2.4 2021 1 OMPA\_RICCN  
35 218.5 2.3 1829 1 FRPC\_NEIMB  
36 218 2.3 1481 1 APU\_THETU  
37 217.5 2.3 1475 1 APU\_THETU  
38 216 2.3 1271 1 Y338\_MICGE  
39 216 2.3 1829 1 FRPC\_NEIMC  
40 213 2.3 828 1 PMFC\_PROMI  
41 210.5 2.2 1541 1 IGA1\_HAEIN  
42 209.5 2.2 1892 1 Y835\_RICCN  
43 207.5 2.2 1508 1 HLVA\_SERMA  
44 206 2.2 1296 1 VACL\_HELPY  
45 206 2.2 1694 1 IGA0\_HAEIN

## ALIGNMENTS

RESULT 1  
GFCD\_STRMU STANDARD; PRT; 1462 AA.  
AC P49331; O69383; O69386; O69389; O69392; O69398;  
-DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
DN (Sucrose 6-glucosyltransferase).  
GN GTFD.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE-91100958; PubMed-2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme."  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN-MT4239, MT4245, MT4251, MT4467, and MT8148;  
RC MEDLINE-98231643; PubMed-9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-  
fructose + ((1,6)-alpha-D-glucosyl)(N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S.MUTANS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M29296; AAA26895.1; -;  
DR EMBL; D88653; AAA26103.1; -;





```
Qy 1580 QVSTGVAIDPSQKITEKSAFYNGTNIHLRGSGVYLKADGGQYVNLGTTTKQFPLQLTG 1639
Db 1027 QISTGVPMDSVKIKQNSAKYFNGTNIHLRGAGVYLKQATNTYFSLVSDNTFLPKSLV- 1085
Qy 1640 EKKGNREGFYKG--NDG-NIYFYDLAGNMVNTPIEDSVG-NNYFFQDQKGMVENKHFDV 1695
Db 1086 NPNHGTSSSVTGLVFDGKGYYVYSTGNAQNAFI--SLGNWYVFNONGYVMTGAOSIN 1143
Qy 1696 VDSYGEKGTFFLKGVSFRGLVQT-----DNQTY-----YFDNYGK 1733
Db 1144 -----GANYFFLNGTQLRNAIYDNGKNVLSYVGNDRRYNGYVLFQGOORYFQX-GI 1196
Qy 1734 MYRNOTINAGAMITLDENGKLIKASYNDAEYPTSTDVGKM 1775
Db 1197 MAVGLTRHGAVOY-FOASG-----FQAKGQFITAD-GKL 1230

RESULT 3
GTFB_STRMU
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, and MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
Db EMBL; M17361; AAA8588.1; -
DR EMBL; D88651; BAA26101.1; -
DR EMBL; D88654; BAA26101.1; -
```

```
DR EMBL; D88657; BAA26109.1; -
DR EMBL; D88660; BAA26113.1; -
DR EMBL; D89977; BAA26119.1; -
DR FIR; B33135; B33135.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF04473; CW_binding_1; 13.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT VARIANT 62 62
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT CONFLICT 1476 AA; 165685 MW; 3479862807694D98 CRC64;
SQ SEQUENCE

Query Match 24.7%; Score 2330.5; DB 1; Length 1476;
Best Local Similarity 42.8%; Pred. No. 2e-100;
Matches 522; Conservative 185; Mismatches 379; Indels 135; Gaps 35;

Qy 637 IVLVNGKEVKROLVND-----TKEGAAGFNNDVYKVNPAINSSM 677
Db 35 LVKADSNESKSIQSDNSNTSVVTANEESNVITEATSKQEAASQTN--HTVTTSSSTSV 92
```





```

SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;
Query Match 24.1%; Score 2274; DB 1; Length 1592;
Best Local Similarity 45.9%; Pred. No. 9e-98;
Matches 496; Conservative 167; Mismatches 340; Indels 78; Gaps 29;

Qy 688 VTKNENVOLVHRSNDVKTEGNYVDFWSELMVPKDSFGKNGPLKQFGLQTINGOQYY 747
Dy 116 VPTENEN-----QGTFDEMLAEAKNVAETASDSIP-----SLAKMSNVKQVDGKYY 163
Qy 748 IDPTTGPRKNFLQSGNNWTFYDSDTGV--GTNALELQPAKGTVSSNEQ--YRNGNAAS 804
Dy 164 YD-QDGNVKNFVSGDKIYFED-ETGAYKDTSKVDADKSSAVSQNATIFRANREAYS 221
Qy 805 YDKSNIENVGLYADTWYRQPKILKCGTTWDSKETDMRPILMVWPNLTQAYLYNYM 864
Dy 222 TSAKNFAVDNLYLADSKNRPKSLKDGKTWETSGDKDFRLLMAWPDFTKRNLYNYM 281
Qy 865 KQHGNLLPSALPFFNADAPAEALNHYSEIVQNIETKRISETGNTDMLRTHLMDHFTVNNPM 924
Dy 282 ----NKVGVIDKTYTAETSQADLTAAAEVLQVARTEQKITSENNTKWLREAISAFVKTQP 337
Qy 925 WNKSENVNFGIOFGQFLKYEN-SLTPYANSDYRLGLRMPINIKDQ-----TYR--- 975
Dy 338 WNGSEKPYDD--HLQNGALLFDNQTDLTPDTQSNRYLLARTPTNQTGSLDSRFTYPNPD 395
Qy 976 ---GOEFLLANDIDNSNPVQAEOLNLYLLNFGTTTANNDOANEDSVRVADPDNIDAD 1032
Dy 396 PLGCVDFLLANDVNSNPVQAEOLNLYLLNFGSIYANDADANFDSIRVADENVDAD 455
Qy 1033 LMNTAODYFNAAYGMD-SDAVSNKHINLELDNHADEYFNKIGNPOLTWDDTIKNSLNH 1091
Dy 456 QLGQSSDYKAAAYGIDKNNKNNHNSIVEAWSNDPTPYLHDDGDNLMNMDNKFRLSMLW 515
Qy 1092 GLSDATN--RWGLDAIHOSLADRENNTENVIPNYFVRAHDNNSODIQONAIR-DVTG 1149
Dy 516 SLAFTDVRSLNLIHNSLVDRVDREVETVPSYFAAHSEVODITRDIKABINP 575
Qy 1150 KDY-HTTFEDEQKIDAYIQDQNSTVKKYNLYNIPASYAILLTNKTIPRVYVGDLYTD 1208
Dy 576 NSFYSFTEQIEIDQAFIYNEDLKKDKKTYHNVPLSYLTLLTNKGSIPRVYVGDYFTD 635
Qy 1209 GGQYMEHOTRYDYLTLNLSRVYVAGGOSMOTMSVGGNNILTSVRYCKGAWATDTG 1268
Dy 636 DGQYANKTVNYDAIESLLKARKYVAGGQAMQYQI--GNGEILTSVRYCKGALKQSDKG 694
Qy 1269 TDETRTOIGVVSNTNLKLVNDKVV-LHMGAAHKNQYRAVLTTDGVINYTSDOG 1327
Dy 695 DATRTSGVGMGNQPNFSL--DCKVVALNMGAAHQAQYRALWSTKDGVAITYATDAD 752
Qy 1328 AP----VAMTDENGDLVLSHNLVYNGKEADTAVQGVANPDYSGYLAVVYPVGCASNDQ 1383
Dy 753 ASKAGLVKRTDENGYLFLNDDL-----KGVANPQVSGFLQVWVPVGAADQDQ 800
Qy 1384 ARTAPSTKNSGNSAYRTNAAFSDSNVIFEAFSNFYVPTKESERANVRIQNADFFASLG 1443
Dy 801 IRVAASDTASTDGKSLHQDAADSRMVFEGFSNFSQSPATKEEYTVNVVIANNVDKFVSWG 860
Qy 1444 FTSFEMAPQYNSKDRFLDSTINGYAFTRDYLGHSENKNGTDDDLRNLQAALHAKG 1503
Dy 861 ITDFEMAPQYVSSYDGFQFQSDVIQNGYAFTRDYLGHSKANKYGTADQLYKATKALHAKG 920
Qy 1504 LQVMADWVPDQIYLPKQKEVATVTRVDRGNVWKDAIINNLYVYNT-IGGGEVQKQYKG 1562
Dy 921 LKVMADWVPDQMYTFPQOEYVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYKG 980
Qy 1563 AFLDLKLYPEITTKQVSTGVDAIDSQKITEWSAKYFNGTNIHLHSGSYVVKAD-GGQ 1621
Dy 981 AFLDLKLYPEITTKQVSTGVDAIDSQKITEWSAKYFNGTNIHLHSGSYVVKAD-GGQ 1621
Qy 1622 YNLGTTTKQPLQLTGKKGQNEGKVGKNDGNYFYFD--LAGNVKNTFIEDSVGNWY 1679
Dy 1041 YFNVASDT-LFLPSLLGKVVESGIRY-----DGKGYINNSATGDQVKASFITEA-ONLY 1094
```

```

Qy 1680 FFDQDKMVENKHFVDVDSYGEKGYFFLKNGVSPRGLVQTD--NGTYFDNYGKMVRN 1737
Dy 1095 YFGKDGVMYTGQNTIN-----GANVFLENCTALR-NIIYTDAGNSHYANDCKRYEN 1147
Qy 1738 Q 1738
Dy 1148 E 1148

RESULT 5
GTF1_STRDO STANDARD; PRT; 1597 AA.
ID GTF1_STRDO
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFI.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RC MEDLINE=87308014; PubMed=3040686;
RA Perretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S.MUTANS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M17391; AAC63063.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 19.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1..38
FT CHAIN 39..1597
FT DOMAIN 39..1050
FT DOMAIN 1050..1597
FT DOMAIN 1099..1597
FT REPEAT 1099..1132
FT REPEAT 1163..1213
FT REPEAT 1227..1277
FT REPEAT 1292..1342
FT REPEAT 1352..1399
FT REPEAT 1406..1455
FT REPEAT 1465..1512
```



```

Db 1044 SNKLVNSDDKLFPLPKYLLQGVESGRFP-----DGTGYVNSSTTGGKVDVSFITEA-GN 1098
QY 1678 WYFFDQDGKVMENKHFVDVDSYGEKGT-YFFLKNVGSFRGLQYOTD--NGTYFYFDNYGKM 1734
Db 1099 LYFEGDQGYMYTGAQNI-----KGSNYVFLANGAALR-NTVYTDAGQGNHHYVYNDGKR 1150
QY 1735 VRN 1737
Db 1151 YEN 1153

RESULT 6
GTF5_STRDO
ID GTF5_STRDO STANDARD; PRT; 1365 AA.
AC P29316;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE Glucosyl transferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTF5.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=13117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RX MEDLINE=90316665; PubMed=2142479;
RA Gilmore K.S., Russell R.R., Ferretti J.J.;
RT Analysis of the Streptococcus downei gtf5 gene, which specifies a
RT glucosyltransferase that synthesizes soluble glucans.";
RL Infect.Immun. 58:2452-2458(1990).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -!- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN UNLIKE GTF-1.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
CC 1,6-GLUCOSE).
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -!- SIMILARITY: CONTAINS 10 CELL WALL BINDING REPEATS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30943; AAA26898.1; -.
CC PIR; A41483; A41483.
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; GH70.
CC Pfam; PF01473; CW_binding_1; 10.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Transferrase; Glucosyltransferase; Signal; Repeat; Dental caries.
CC FT SIGNAL 1 36 OR 37 (POTENTIAL).
CC FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
CC FT REPEAT 157 177 CELL WALL BINDING 1.
CC FT REPEAT 178 197 CELL WALL BINDING 2.
CC FT DOMAIN 198 1061 CATALYTIC (APPROXIMATE).
CC FT REPEAT 1062 1082 CELL WALL BINDING 3.
CC FT REPEAT 1083 1102 CELL WALL BINDING 4.
CC FT REPEAT 1150 1169 CELL WALL BINDING 5.
CC FT REPEAT 1170 1190 CELL WALL BINDING 6.

```

```
FT REPEAT 1225 1243 CELL WALL BINDING 7.
FT REPEAT 1289 1308 CELL WALL BINDING 8.
FT REPEAT 1309 1328 CELL WALL BINDING 9.
FT REPEAT 1331 1352 CELL WALL BINDING 10.
SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 23.3%; Score 2191.5; DB 1; Length 1365;
Best Local Similarity 40.9%; Pred. Mismatches 436; Indels 141; Gaps 34;
Matches 518; Conservative 170;

OY 581 AADGNPDGNTVDTQFSKNYATGNGFDYVKVNGNOVFSGWHATQNSDKDSQWIVLV 640
DB 36 AADTNNDGTSV--QVVK-MVPSDPKFDQAQNG-QLAQAFKAANAQDQATATSOVS PAT 91
OY 641 NGREVKPOLYNDYKAGAGRNNDYKYNPAIENSNSGFGIITLPTVYKNEWQVLVHR 700
DB 92 DGR-VDMQVTPAANQPAANVANDV--ANPATDAGAL-----NROSAAD 132
OY 701 FSDVKTGEGNYDFWSELMPVKDSFGKNGPLKQFGLQTINGQQYIDPTTQGPKNFL 760
DB 133 TSTDGKA-----VP-QTSDQPGH-----LEIVDGKTYVD-ANGORLKNYS 171
OY 761 LOSGNWYFDSGTGVTNLALELOFAKGTVSSNEQYRNGNAAYSDDKSIENTVNGILTAD 820
DB 172 MVIDGKTYFDGQTGEAQTLPKLTQANQNDVPDPSQANNOAYSNEASSPETYDNYLTAD 231
OY 821 TWYRPKQILKDGTTWDSKETDRPILMWMPNTLTQAYLYNMYKOHGNLPSALPFENA 880
DB 232 SWYRPKILNKGQSQWASSEGDLRPILMTWMPDAATKAAAYANFWAKEG-LISGSYQNSA 290
OY 881 DADPAELNHSEIVQONIEKRISETGNTDMLRLMHDFTVNNPMWKNDSNVNFSGIQ-- 938
DB 291 NLDAATON----TQSAIERKKIASSEGTNWLRLDKMSQFVKSQNSQWASSTASNETVY PQNH 345
OY 939 FQGGFLAYKNSDLTPYANSYRILGRMPI-----NIKDTYRGOEFLLANDINDSNPVV 992
DB 346 MQGALLFSNKSDEHANSQWRLNRPFTQTKQKQYFTTYNAGYELLANDVDSNPVV 405
OY 993 QAEQLNLWLYLLNGTITANNQANFDSVRVDPADNTDADLMNTAQDYFNAAYCMD-SDA 1051
DB 406 QAEQLNLHLYLMNGDITVMDGDKANFVGDRVDAVDNYNADLLQTDYRYKAKYTDQNEK 465
OY 1052 VSNKHINILEDWNADEPEYENKIGNPOLTMDPTIKNSLNLGLSDAT-NRWGLDAIVHOSL 1110
DB 466 NAIDHLSILEAWSGNDYKQDNFSLSDNDQSRMSLKAFGVASYRNLNLSLATAGL 525
OY 1111 ADRENNSTENVIPNYSEVRAHDKNSODQIONAIRDVTGK----DYHTFTFEDEQKGIDA 1166
DB 526 KNRSANPDSPV-PNYVFIKRAHSEVQTRIAKIREKLGKTNADGLTNLTDLDLNAFADI 584
OY 1167 YIQDQNSVKKYLYNIPASVAILLTNKTIPRYVYIGDLYTDGGOYMEHQTRYDITLNL 1226
DB 585 YNQDMNAPDKVYYPNNLPMAFAMWLQNKDITVRYVYIGDLYTDGGOYMATKTPFYNAIETL 644
OY 1227 LKSRVYVAGQSM---QTMVSVGNNILTSVRVYKGMATATDGTDTQIGVYVVSN 1283
DB 645 LKGRIVYVAGQVSYKQDRSSG---ILTSVRVYKGMATATDGTDTQIGVYVVSN 1283
OY 1284 TPNKLGVDKXVVLHMGAAHKNQYQRAVLTITDGVNTYSD---QCAPVMTDEGNDLY 1340
DB 701 RPNFRAYRN--LTLNMGAAHKSQAYRELLSTKDGIAITYLNDSDVDSDQYKYKDSQGNLS 758
OY 1341 LSSHLVYNGKEEADTAQGVANPDVSGYLAVVYVPCASDQDARTAPSTEKSGNSAYR 1400
DB 759 FSASEL-----QSVANQVSGMTQVWVPVGAADQDVRYSFQATKDGNIYH 806
OY 1401 TNAFDSNVIFEAFSNFYPTPKESRANVRIAONADFFASLGFTSFEMAPOYNSSKDR 1460
DB 807 QSDALDSQVIEGFSNFAQFASQPDQVNTAVIAKNGDLFSWGTQPEMAPQVYVSSDGT 866
OY 1461 FLDSTIDNGYAFTRDYLGLSEPNKYCTDDELRNAIQALHKAGLOYWADVVPQIYNLPG 1520
DB 867 FLDSVILNGYAFSDRYDLAMSNNKNGYKSKQDLANAIKGLQSAGIKVLSLDLPVNPQYLN 926
```

## RESULT 7

```
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR NI7G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the hic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacXY region."
RL Microbiology 142:3113-3123(1996).
CC CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC CC INTO THE MEDIUM.
CC CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE
CC CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
```

CC MOTIF REPEATED 31 TIMES.  
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME  
CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
CC -----  
CC EMBL: L05634; AAA22883.1; -  
CC EMBL: D31856; BAA06656.1; -  
CC EMBL: D29985; BAA06260.1; -  
CC EMBL: D83026; BAA11683.1; -  
CC EMBL: Z99124; CABI5959.1; -  
CC PIR: S32920; S32920.  
CC Subtilisin; BG10797; wapa.  
CC InterPro: IPR003305; CBM\_CenC.  
CC Pfam: PF02018; CBM\_4\_9; 1.  
CC Cell wall; Repeat; Signal; Complete proteome.  
CC SIGNAL 1 28  
CC CHAIN 29 2334  
CC DOMAIN 504 869  
CC REPEAT 504 605  
CC REPEAT 636 736  
CC REPEAT 769 869  
CC DOMAIN 1021 2139  
CC 31 X 21 AA APPROXIMATE TANDEM REPEATS OF  
CC X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).  
CC 2-1. 1040  
CC REPEAT 1042 1061  
CC REPEAT 1063 1082  
CC REPEAT 1083 1102  
CC REPEAT 1109 1128  
CC REPEAT 1129 1148  
CC REPEAT 1150 1169  
CC REPEAT 1174 1193  
CC REPEAT 1199 1218  
CC REPEAT 1219 1238  
CC REPEAT 1246 1265  
CC REPEAT 1267 1286  
CC REPEAT 1287 1306  
CC REPEAT 1307 1326  
CC REPEAT 1327 1346  
CC REPEAT 1347 1366  
CC REPEAT 1367 1386  
CC REPEAT 1387 1405  
CC REPEAT 1406 1425  
CC REPEAT 1426 1445  
CC REPEAT 1446 1464  
CC REPEAT 1465 1483  
CC REPEAT 1484 1502  
CC REPEAT 1503 1521  
CC REPEAT 1522 1540  
CC REPEAT 1541 1559  
CC REPEAT 1560 1578  
CC REPEAT 1579 1597  
CC REPEAT 1598 1616  
CC REPEAT 1617 1635  
CC REPEAT 1636 1654  
CC REPEAT 1655 1673  
CC REPEAT 1674 1692  
CC REPEAT 1693 1711  
CC REPEAT 1712 1730  
CC REPEAT 1731 1749  
CC REPEAT 1750 1768  
CC REPEAT 1769 1787  
CC REPEAT 1788 1806  
CC REPEAT 1807 1825  
CC REPEAT 1826 1844  
CC REPEAT 1845 1862  
CC REPEAT 1863 1881  
CC REPEAT 1882 1900  
CC REPEAT 1901 1919  
CC REPEAT 1920 1938  
CC REPEAT 1939 1957  
CC REPEAT 1958 1976  
CC REPEAT 1977 1995  
CC REPEAT 1996 2014  
CC REPEAT 2015 2033  
CC REPEAT 2034 2052  
CC REPEAT 2053 2070  
CC REPEAT 2071 2089  
CC REPEAT 2090 2108  
CC REPEAT 2109 2127  
CC REPEAT 2128 2146  
CC REPEAT 2147 2165  
CC REPEAT 2166 2184  
CC REPEAT 2185 2203  
CC REPEAT 2204 2222  
CC REPEAT 2223 2241  
CC REPEAT 2242 2260  
CC REPEAT 2261 2279  
CC REPEAT 2280 2298  
CC REPEAT 2299 2317  
CC REPEAT 2318 2336  
CC REPEAT 2337 2355  
CC REPEAT 2356 2374  
CC REPEAT 2375 2393  
CC REPEAT 2394 2412  
CC REPEAT 2413 2431  
CC REPEAT 2432 2450  
CC REPEAT 2451 2469  
CC REPEAT 2470 2488  
CC REPEAT 2489 2507  
CC REPEAT 2508 2526  
CC REPEAT 2527 2545  
CC REPEAT 2546 2564  
CC REPEAT 2565 2583  
CC REPEAT 2584 2602  
CC REPEAT 2603 2621  
CC REPEAT 2622 2640  
CC REPEAT 2641 2659  
CC REPEAT 2660 2678  
CC REPEAT 2679 2697  
CC REPEAT 2698 2716  
CC REPEAT 2717 2735  
CC REPEAT 2736 2754  
CC REPEAT 2755 2773  
CC REPEAT 2774 2792  
CC REPEAT 2793 2811  
CC REPEAT 2812 2830  
CC REPEAT 2831 2849  
CC REPEAT 2850 2868  
CC REPEAT 2869 2887  
CC REPEAT 2888 2906  
CC REPEAT 2907 2925  
CC REPEAT 2926 2944  
CC REPEAT 2945 2963  
CC REPEAT 2964 2982  
CC REPEAT 2983 3001  
CC REPEAT 3002 3020  
CC REPEAT 3021 3039  
CC REPEAT 3040 3058  
CC REPEAT 3059 3077  
CC REPEAT 3078 3096  
CC REPEAT 3097 3115  
CC REPEAT 3116 3134  
CC REPEAT 3135 3153  
CC REPEAT 3154 3172  
CC REPEAT 3173 3191  
CC REPEAT 3192 3210  
CC REPEAT 3211 3229  
CC REPEAT 3230 3248  
CC REPEAT 3249 3267  
CC REPEAT 3268 3286  
CC REPEAT 3287 3305  
CC REPEAT 3306 3324  
CC REPEAT 3325 3343  
CC REPEAT 3344 3362  
CC REPEAT 3363 3381  
CC REPEAT 3382 3400  
CC REPEAT 3401 3419  
CC REPEAT 3420 3438  
CC REPEAT 3439 3457  
CC REPEAT 3458 3476  
CC REPEAT 3477 3495  
CC REPEAT 3496 3514  
CC REPEAT 3515 3533  
CC REPEAT 3534 3552  
CC REPEAT 3553 3571  
CC REPEAT 3572 3590  
CC REPEAT 3591 3609  
CC REPEAT 3610 3628  
CC REPEAT 3629 3647  
CC REPEAT 3648 3666  
CC REPEAT 3667 3685  
CC REPEAT 3686 3704  
CC REPEAT 3705 3723  
CC REPEAT 3724 3742  
CC REPEAT 3743 3761  
CC REPEAT 3762 3780  
CC REPEAT 3781 3799  
CC REPEAT 3799 3818  
CC REPEAT 3819 3837  
CC REPEAT 3838 3856  
CC REPEAT 3857 3875  
CC REPEAT 3876 3894  
CC REPEAT 3895 3913  
CC REPEAT 3914 3932  
CC REPEAT 3933 3951  
CC REPEAT 3952 3970  
CC REPEAT 3971 3989  
CC REPEAT 3990 4008  
CC REPEAT 4009 4027  
CC REPEAT 4028 4046  
CC REPEAT 4047 4065  
CC REPEAT 4066 4084  
CC REPEAT 4085 4103  
CC REPEAT 4104 4122  
CC REPEAT 4123 4141  
CC REPEAT 4142 4160  
CC REPEAT 4161 4179  
CC REPEAT 4180 4198  
CC REPEAT 4199 4217  
CC REPEAT 4218 4236  
CC REPEAT 4237 4255  
CC REPEAT 4256 4274  
CC REPEAT 4275 4293  
CC REPEAT 4294 4312  
CC REPEAT 4313 4331  
CC REPEAT 4332 4350  
CC REPEAT 4351 4369  
CC REPEAT 4370 4388  
CC REPEAT 4389 4407  
CC REPEAT 4408 4426  
CC REPEAT 4427 4445  
CC REPEAT 4446 4464  
CC REPEAT 4465 4483  
CC REPEAT 4484 4502  
CC REPEAT 4503 4521  
CC REPEAT 4522 4540  
CC REPEAT 4541 4559  
CC REPEAT 4560 4578  
CC REPEAT 4579 4597  
CC REPEAT 4598 4616  
CC REPEAT 4617 4635  
CC REPEAT 4636 4654  
CC REPEAT 4655 4673  
CC REPEAT 4674 4692  
CC REPEAT 4693 4711  
CC REPEAT 4712 4730  
CC REPEAT 4731 4749  
CC REPEAT 4750 4768  
CC REPEAT 4769 4787  
CC REPEAT 4788 4806  
CC REPEAT 4807 4825  
CC REPEAT 4826 4844  
CC REPEAT 4845 4863  
CC REPEAT 4864 4882  
CC REPEAT 4883 4901  
CC REPEAT 4902 4920  
CC REPEAT 4921 4939  
CC REPEAT 4940 4958  
CC REPEAT 4959 4977  
CC REPEAT 4978 4996  
CC REPEAT 4997 5015  
CC REPEAT 5016 5034  
CC REPEAT 5035 5053  
CC REPEAT 5054 5072  
CC REPEAT 5073 5091  
CC REPEAT 5092 5110  
CC REPEAT 5111 5129  
CC REPEAT 5130 5148  
CC REPEAT 5149 5167  
CC REPEAT 5168 5186  
CC REPEAT 5187 5205  
CC REPEAT 5206 5224  
CC REPEAT 5225 5243  
CC REPEAT 5244 5262  
CC REPEAT 5263 5281  
CC REPEAT 5282 5300  
CC REPEAT 5301 5319  
CC REPEAT 5320 5338  
CC REPEAT 5339 5357  
CC REPEAT 5358 5376  
CC REPEAT 5377 5395  
CC REPEAT 5396 5414  
CC REPEAT 5415 5433  
CC REPEAT 5434 5452  
CC REPEAT 5453 5471  
CC REPEAT 5472 5490  
CC REPEAT 5491 5509  
CC REPEAT 5510 5528  
CC REPEAT 5529 5547  
CC REPEAT 5548 5566  
CC REPEAT 5567 5585  
CC REPEAT 5586 5604  
CC REPEAT 5605 5623  
CC REPEAT 5624 5642  
CC REPEAT 5643 5661  
CC REPEAT 5662 5680  
CC REPEAT 5681 5699  
CC REPEAT 5699 5718  
CC REPEAT 5719 5737  
CC REPEAT 5738 5756  
CC REPEAT 5757 5775  
CC REPEAT 5776 5794  
CC REPEAT 5795 5813  
CC REPEAT 5814 5832  
CC REPEAT 5833 5851  
CC REPEAT 5852 5869  
CC REPEAT 5870 5889  
CC REPEAT 5889 5908  
CC REPEAT 5909 5927  
CC REPEAT 5928 5946  
CC REPEAT 5947 5965  
CC REPEAT 5966 5984  
CC REPEAT 5985 6003  
CC REPEAT 6004 6022  
CC REPEAT 6023 6041  
CC REPEAT 6042 6060  
CC REPEAT 6061 6079  
CC REPEAT 6080 6098  
CC REPEAT 6099 6117  
CC REPEAT 6118 6136  
CC REPEAT 6137 6155  
CC REPEAT 6156 6174  
CC REPEAT 6175 6193  
CC REPEAT 6194 6212  
CC REPEAT 6213 6231  
CC REPEAT 6232 6250  
CC REPEAT 6251 6269  
CC REPEAT 6270 6288  
CC REPEAT 6289 6307  
CC REPEAT 6308 6326  
CC REPEAT 6327 6345  
CC REPEAT 6346 6364  
CC REPEAT 6365 6383  
CC REPEAT 6384 6402  
CC REPEAT 6403 6421  
CC REPEAT 6422 6440  
CC REPEAT 6441 6459  
CC REPEAT 6460 6478  
CC REPEAT 6479 6497  
CC REPEAT 6498 6516  
CC REPEAT 6517 6535  
CC REPEAT 6536 6554  
CC REPEAT 6555 6573  
CC REPEAT 6574 6592  
CC REPEAT 6593 6611  
CC REPEAT 6612 6630  
CC REPEAT 6631 6649  
CC REPEAT 6650 6668  
CC REPEAT 6669 6687  
CC REPEAT 6688 6706  
CC REPEAT 6707 6725  
CC REPEAT 6726 6744  
CC REPEAT 6745 6763  
CC REPEAT 6764 6782  
CC REPEAT 6783 6801  
CC REPEAT 6802 6820  
CC REPEAT 6821 6839  
CC REPEAT 6840 6858  
CC REPEAT 6859 6877  
CC REPEAT 6878 6896  
CC REPEAT 6897 6915  
CC REPEAT 6916 6934  
CC REPEAT 6935 6953  
CC REPEAT 6954 6972  
CC REPEAT 6973 6991  
CC REPEAT 6992 7009  
CC REPEAT 7010 7028  
CC REPEAT 7029 7047  
CC REPEAT 7048 7066  
CC REPEAT 7067 7085  
CC REPEAT 7086 7104  
CC REPEAT 7105 7123  
CC REPEAT 7124 7142  
CC REPEAT 7143 7161  
CC REPEAT 7162 7179  
CC REPEAT 7180 7200  
CC REPEAT 7201 7220  
CC REPEAT 7221 7239  
CC REPEAT 7240 7258  
CC REPEAT 7259 7277  
CC REPEAT 7278 7296  
CC REPEAT 7297 7315  
CC REPEAT 7316 7334  
CC REPEAT 7335 7353  
CC REPEAT 7354 7372  
CC REPEAT 7373 7391  
CC REPEAT 7392 7410  
CC REPEAT 7411 7429  
CC REPEAT 7430 7448  
CC REPEAT 7449 7467  
CC REPEAT 7468 7486  
CC REPEAT 7487 7505  
CC REPEAT 7506 7524  
CC REPEAT 7525 7543  
CC REPEAT 7544 7562  
CC REPEAT 7563 7581  
CC REPEAT 7582 7600  
CC REPEAT 7601 7619  
CC REPEAT 7620 7638  
CC REPEAT 7639 7657  
CC REPEAT 7658 7676  
CC REPEAT 7677 7695  
CC REPEAT 7696 7714  
CC REPEAT 7715 7733  
CC REPEAT 7734 7752  
CC REPEAT 7753 7771  
CC REPEAT 7772 7790  
CC REPEAT 7791 7809  
CC REPEAT 7810 7828  
CC REPEAT 7829 7847  
CC REPEAT 7848 7866  
CC REPEAT 7867 7885  
CC REPEAT 7886 7904  
CC REPEAT 7905 7923  
CC REPEAT 7924 7942  
CC REPEAT 7943 7961  
CC REPEAT 7962 7980  
CC REPEAT 7981 7999  
CC REPEAT 7999 8018  
CC REPEAT 8019 8037  
CC REPEAT 8038 8056  
CC REPEAT 8057 8075  
CC REPEAT 8076 8094  
CC REPEAT 8095 8113  
CC REPEAT 8114 8132  
CC REPEAT 8133 8151  
CC REPEAT 8152 8169  
CC REPEAT 8170 8188  
CC REPEAT 8189 8207  
CC REPEAT 8208 8226  
CC REPEAT 8227 8245  
CC REPEAT 8246 8264  
CC REPEAT 8265 8283  
CC REPEAT 8284 8302  
CC REPEAT 8303 8321  
CC REPEAT 8322 8340  
CC REPEAT 8341 8359  
CC REPEAT 8360 8378  
CC REPEAT 8379 8397  
CC REPEAT 8398 8416  
CC REPEAT 8417 8435  
CC REPEAT 8436 8454  
CC REPEAT 8455 8473  
CC REPEAT 8474 8492  
CC REPEAT 8493 8511  
CC REPEAT 8512 8529  
CC REPEAT 8530 8548  
CC REPEAT 8549 8567  
CC REPEAT 8568 8586  
CC REPEAT 8587 8605  
CC REPEAT 8606 8624  
CC REPEAT 8625 8643  
CC REPEAT 8644 8662  
CC REPEAT 8663 8681  
CC REPEAT 8682 8700  
CC REPEAT 8701 8719  
CC REPEAT 8720 8738  
CC REPEAT 8739 8757  
CC REPEAT 8758 8776  
CC REPEAT 8777 8795  
CC REPEAT 8796 8814  
CC REPEAT 8815 8833  
CC REPEAT 8834 8852  
CC REPEAT 8853 8871  
CC REPEAT 8872 8890  
CC REPEAT 8891 8909  
CC REPEAT 8910 8928  
CC REPEAT 8929 8947  
CC REPEAT 8948 8966  
CC REPEAT 8967 8985  
CC REPEAT 8986 9004  
CC REPEAT 9005 9023  
CC REPEAT 9024 9042  
CC REPEAT 9043 9061  
CC REPEAT 9062 9080  
CC REPEAT 9081 9099  
CC REPEAT 9100 9118  
CC REPEAT 9119 9137  
CC REPEAT 9138 9156  
CC REPEAT 9157 9175  
CC REPEAT 9176 9194  
CC REPEAT 9195 9213  
CC REPEAT 9214 9232  
CC REPEAT 9233 9251  
CC REPEAT 9252 9269  
CC REPEAT 9270 9288  
CC REPEAT 9289 9307  
CC REPEAT 9308 9326  
CC REPEAT 9327 9345  
CC REPEAT 9346 9364  
CC REPEAT 9365 9383  
CC REPEAT 9384 9402  
CC REPEAT 9403 9421  
CC REPEAT 9422 9439  
CC REPEAT 9440 9458  
CC REPEAT 9459 9477  
CC REPEAT 9478 9496  
CC REPEAT 9497 9515  
CC REPEAT 9516 9534  
CC REPEAT 9535 9553  
CC REPEAT 9554 9572  
CC REPEAT 9573 9591  
CC REPEAT 9592 9610  
CC REPEAT 9611 9629  
CC REPEAT 9630 9648  
CC REPEAT 9649 9667  
CC REPEAT 9668 9686  
CC REPEAT 9687 9705  
CC REPEAT 9706 9724  
CC REPEAT 9725 9743  
CC REPEAT 9744 9762  
CC REPEAT 9763 9781  
CC REPEAT 9782 9800  
CC REPEAT 9801 9819  
CC REPEAT 9820 9838  
CC REPEAT 9839 9857  
CC REPEAT 9858 9876  
CC REPEAT 9877 9895  
CC REPEAT 9896 9914  
CC REPEAT 9915 9933  
CC REPEAT 9934 9952  
CC REPEAT 9953 9971  
CC REPEAT 9972 9990  
CC REPEAT 9991 10009  
CC REPEAT 10010 10029  
CC REPEAT 10030 10048  
CC REPEAT 10049 10067  
CC REPEAT 10068 10086  
CC REPEAT 10087 10105  
CC REPEAT 10106 10124  
CC REPEAT 10125 10143  
CC REPEAT 10144 10162  
CC REPEAT 10163 10181  
CC REPEAT 10182 10200  
CC REPEAT 10201 10219  
CC REPEAT 10220 10238  
CC REPEAT 10239 10257  
CC REPEAT 10258 10276  
CC REPEAT 10277 10295  
CC REPEAT 10296 10314  
CC REPEAT 10315 10333  
CC REPEAT 10334 10352  
CC REPEAT 10353 10371  
CC REPEAT 10372 10390  
CC REPEAT 10391 10409  
CC REPEAT 10410 10428  
CC REPEAT 10429 10447  
CC REPEAT 10448 10466  
CC REPEAT 10467 10485  
CC REPEAT 10486 10504  
CC REPEAT 10505 10523  
CC REPEAT 10524 10542  
CC REPEAT 10543 10561  
CC REPEAT 10562 10579  
CC REPEAT 10580 10600  
CC REPEAT 10601 10620  
CC REPEAT 10621 10640  
CC REPEAT 10641 10660  
CC REPEAT 10661 10680  
CC REPEAT 10681 10700  
CC REPEAT 10701 10720  
CC REPEAT 10721 10740  
CC REPEAT 10741 10760  
CC REPEAT 10761 10780  
CC REPEAT 10781 10800  
CC REPEAT 10801 10820  
CC REPEAT 10821 10840  
CC REPEAT 10841 10860  
CC REPEAT 10861 10880  
CC REPEAT 10881 10900  
CC REPEAT 10901 10920  
CC REPEAT 10921 10940  
CC REPEAT 10941 10960  
CC REPEAT 10961 10980  
CC REPEAT 10981 11000  
CC REPEAT 11001 11020  
CC REPEAT 11021 11040  
CC REPEAT 11041 11060  
CC REPEAT 11061 11080  
CC REPEAT 11081 11100  
CC REPEAT 11101 11120  
CC REPEAT 11121 11140  
CC REPEAT 11141 11160  
CC REPEAT 11161 11180  
CC REPEAT 11181 11200  
CC REPEAT 11201 11220  
CC REPEAT 11221 11240  
CC REPEAT 11241 11260  
CC REPEAT 11261 11280  
CC REPEAT 11281 11300  
CC REPEAT 11301 11320  
CC REPEAT 11321 11340  
CC REPEAT 11341 11360  
CC REPEAT 11361 11380  
CC REPEAT 11381 11400  
CC REPEAT 11401 11420  
CC REPEAT 11421 11440  
CC REPEAT 11441 11460  
CC REPEAT 11461 11480  
CC REPEAT 11481 11500  
CC REPEAT 11501 11520  
CC REPEAT 11521 11540  
CC REPEAT 11541 11560  
CC REPEAT 11561 11580  
CC REPEAT 11581 11600  
CC REPEAT 11601 11620  
CC REPEAT 11621 11640  
CC REPEAT 11641 11660  
CC REPEAT 11661 11680  
CC REPEAT 11681 11700  
CC REPEAT 11701 11720  
CC REPEAT 11721 11740  
CC REPEAT 11741 11760  
CC REPEAT 11761 11780  
CC REPEAT 11781 11800  
CC REPEAT 11801 11820  
CC REPEAT 11821 11840  
CC REPEAT 11841 11860  
CC REPEAT 11861 11880  
CC REPEAT 11881 11900  
CC REPEAT 11901 11920  
CC REPEAT 11921 11940  
CC REPEAT 11941 11960  
CC REPEAT 11961 11980  
CC REPEAT 11981 12000  
CC REPEAT 12001 12020  
CC REPEAT 12021 12040  
CC REPEAT 12041 12060  
CC REPEAT 12061 12080  
CC REPEAT 12081 12100  
CC REPEAT 12101 12120  
CC REPEAT 12121 12140  
CC REPEAT 12141 12160  
CC REPEAT 12161 12180  
CC REPEAT 12181 12200  
CC REPEAT 12201 12220  
CC REPEAT 12221 12240  
CC REPEAT 12241 12260  
CC REPEAT 12261 12280  
CC REPEAT 12281 12300  
CC REPEAT 12301 12320  
CC REPEAT 12321 12340  
CC REPEAT 12341 12360  
CC REPEAT 12361 12380  
CC REPEAT 12381 12400  
CC REPEAT 12401 12420  
CC REPEAT 12421 12440  
CC REPEAT 12441 12460  
CC REPEAT 12461 12480  
CC REPEAT 12481 12500  
CC REPEAT 12501 12520  
CC REPEAT 12521 12540  
CC REPEAT 12541 12560  
CC REPEAT 12561 12580  
CC REPEAT 12581 12600  
CC REPEAT 12601 12620  
CC REPEAT 12621 12640  
CC REPEAT 12641 12660  
CC REPEAT 12661 12680  
CC REPEAT 12681 12700  
CC REPEAT 12701 12720  
CC REPEAT 12721 12740  
CC REPEAT 12741 12760  
CC REPEAT 12761 12780  
CC REPEAT 12781 12800  
CC REPEAT 12801 12820  
CC REPEAT 12821 12840  
CC REPEAT 12841 12860  
CC REPEAT 12861 12880  
CC REPEAT 12881 12900  
CC REPEAT 12901 12920  
CC REPEAT 12921 12940  
CC REPEAT 12941 12960  
CC REPEAT 12961 12980  
CC REPEAT 12981 13000  
CC REPEAT 13001 13020  
CC REPEAT 13021 13040  
CC REPEAT 13041 13060  
CC REPEAT 13061 13080  
CC REPEAT 13081 13100  
CC REPEAT 13101 13120  
CC REPEAT 13121 13140  
CC REPEAT 13141 13160  
CC REPEAT 13161 13180  
CC REPEAT 13181 13200  
CC REPEAT 13201 13220  
CC REPEAT 13221 13240  
CC REPEAT 13241 13260  
CC REPEAT 13261 13280  
CC REPEAT 13281 13300  
CC REPEAT 13301 13320  
CC REPEAT 13321 13340  
CC REPEAT 13341 13360  
CC REPEAT 13361 13380  
CC REPEAT 13381 13400  
CC REPEAT 13401 13420  
CC REPEAT 13421 13440  
CC REPEAT 13441 13460  
CC REPEAT 13461 13480  
CC REPEAT 13481 13500  
CC REPEAT 13501 13520  
CC REPEAT 13521 13540  
CC REPEAT 13541 13560  
CC REPEAT 13561 13580  
CC REPEAT 13581 13600  
CC REPEAT 13601 13620  
CC REPEAT 13621 13640  
CC REPEAT 13641 13660  
CC REPEAT 13661 13680  
CC REPEAT 13681 13700  
CC REPEAT 13701 13720  
CC REPEAT 13721 13740  
CC REPEAT 13741 13760  
CC REPEAT 13761 13780  
CC REPEAT 13781 13800  
CC REPEAT 13801 13820  
CC REPEAT 13821 13840  
CC REPEAT 13841 13860  
CC REPEAT 13861 13880  
CC REPEAT 13881 13900  
CC REPEAT 13901 13920  
CC REPEAT 13921 13940  
CC REPEAT 13941 13960  
CC REPEAT 13961 13980  
CC REPEAT 13981 14000  
CC REPEAT 14001 14020  
CC REPEAT 14021 14040  
CC REPEAT 14041 14060  
CC REPEAT 14061 14080  
CC REPEAT 14081 14100  
CC REPEAT 14101 14120  
CC REPEAT 14121 14140  
CC REPEAT 14141 14160  
CC REPEAT 14161 14180  
CC REPEAT 14181 14200  
CC REPEAT 14201 14220  
CC REPEAT 14221 14240  
CC REPEAT 14241 14260  
CC REPEAT 14261 14280  
CC REPEAT 14281 14300  
CC REPEAT 14301 14320  
CC REPEAT 14321 14340  
CC REPEAT 14341 14360  
CC REPEAT 14361 14380  
CC REPEAT 14381 14400  
CC REPEAT 14401 14420  
CC REPEAT 14421 14440  
CC REPEAT 14441 14460  
CC REPEAT 14461 14480  
CC REPEAT 14481 14500  
CC REPEAT 14501 14520  
CC REPEAT 14521 14540  
CC REPEAT 14541 14560  
CC REPEAT 14561 14580  
CC REPEAT 14581 14600  
CC REPEAT 14601 14620  
CC REPEAT 14621 14640  
CC REPEAT 14641 14660  
CC REPEAT 14661 14680  
CC REPEAT 14681 14700  
CC REPEAT 14701 14720  
CC REPEAT 14721 14740  
CC REPEAT 14741 14760  
CC REPEAT 14761 14780  
CC REPEAT 14781 14800  
CC REPEAT 14801 14820  
CC REPEAT 14821 14840  
CC REPEAT 14841 14860  
CC REPEAT 14861 14880  
CC REPEAT 14881 14900  
CC REPEAT 14901 14920  
CC REPEAT 14921 14940  
CC REPEAT 14941 14960  
CC REPEAT 14961 14980  
CC REPEAT 14981 15000  
CC REPEAT 15001 15020  
CC REPEAT 15021 15040  
CC REPEAT 15041 15060  
CC REPEAT 15061 15080  
CC REPEAT 15081 15100  
CC REPEAT 15101 15120  
CC REPEAT 15121 15140  
CC REPEAT 15141 15160  
CC REPEAT 15161 15180  
CC REPEAT 15181 15200  
CC REPEAT 15201 15220  
CC REPEAT 15221 15240  
CC REPEAT 15241 15260  
CC REPEAT 15261 15280  
CC REPEAT 15281 15300  
CC REPEAT 15301 15320  
CC REPEAT 15321 15340  
CC REPEAT 15

Qy	846	ILMWWNTLTOAYLNMKOHGNLLPSALP-----FFNADADPAELNHYSEIVQON--	897
Db	1101	VKQYSAKSTKAPVTEY-QTSGRRLKAKAINAKETVYSIDADKKTL-----LMTQNGR	1155
Qy	898	-IEKRISSETGN-----TDLRL-TLMHDFVTNNPMNKDSNV-----N	933
Db	1156	KVQYGYNEAGNPIQVIDDAEGLKITTNTKYEKNVVEDPNDVGTGKATESYQYDKDGN	1215
Qy	934	FSGTOFOGGFLKYE---NSDLTPVAN-----SDYRLLGRMPINTKDQTYR-----	975
Db	1216	VTSVKDAYGTETYENKNNVDTKWKDTEGNVTDIAYDGLDAVSETDOSGSSAAVYDKY	1275
Qy	976	GOEFLANDIDNSPVV-----QAEQLNWLXYLLNFGTITANNDQANFDSV-----	1021
Db	1276	GNQIOSSKDLASATNLDKGSFEAKGSGW-----NLTAASKDRRKISVIADKSGVLSG	1327
Qy	1022	-----RVADPN-----IDADLMNTAQDYFNNAAGMDS	1049
Db	1328	SKALEVLSQSTASGTHGYSSATQVLEPNNTVTLGKTKIDIAK-SRAYFNIDL-RDK	1385
Qy	1050	DA-----VSNKHIL-EDW-----NHADPE-----YFNKIG	1075
Db	1386	DOKRIOWIHNEYSALAGKNDWTKRQITFTTPANAGKAVVMEVDKDKDGKAWFEVQ	1445
Qy	1076	NPOLTMDDTIKNSLNHGLSDATNRWGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNN	1135
Db	1446	LEKEVSSSYNPVONSFTSATENW---VSGASVDSSEGFNDV-----SLKAARTSA	1496
Qy	1136	SO-----DOION-----AIRDVTGKDXH-----TFTEDEKKGIDA	1166
Db	1497	SOAGSVTKQTVVLQGSANDKPVYTLTGMSKASSVKTKDEKDSLQANVIYADGSTGI--	1554
Qy	1167	YIDQNSVTKKYN-LYNIPASYAI-----LJTNKOTIPRVYVGLDLYDGGQYMEHQTR	1218
Db	1555	YNAXFPSPQTOENWAAVVPKTKPKNKVD-SILPKSAGTGVVPPDDIRLIEGSLTKST-	1613
Qy	1219	YDILTLLKSRVYVAGGSMQTSVGGNNILTSVRYKGMTATD---TG--TDETR	1273
Db	1614	-YDS-----NGNYVTKDEELGVAITSYDYDETKKTTSETD	1647
Qy	1274	TQIGVYVVSNTPNLKLGVNDRKVLHMGAAHKNQOYRAAVLTDTGVINYTSDDGAPVA-M	1332
Db	1648	AKG-----EKT-----TYTYDQADQLTNM	1666
Qy	1333	TDENGDLYLSSHNLVVGKEADT---AVQGYA-NPDVSGYLAVWV-PVG---ASNQD	1383
Db	1667	TLNGTSLHSYD---KEGNEVSKTIRAGADQTYKFEYDVMGKLVKTTDPLGNVLASEYDA	1724
Qy	1384	ARTAPSTEKNSGNSAYRTNAFD---SNVIFEAFSNFVYTPTKESERANVRIAQNADPFAS	1441
Db	1725	NSNLTKTISPNGNEVLSYDGTDRVKSXYNGTEKY:FTYDKNGNETSV-----	1773
Qy	1442	LGFTSFEMAPQYNSKSKRTFLDSTIDNGYAPTDRYDLQMSKPNKYGTDEDLRNAIQLH-	1500
Db	1774	-----VNKEQNTTKRTE-----DNKNRLTELTDRGSGQTWTPSDSKLTFWSIHG	1821
Qy	1501	-KAGLQVMADWVPOIYNLPKGEVATVTRVDRGNVVK-----	1537
Db	1822	DQKGTNQFTYKLDQMIEMKDSSTSYSEFDYDENGENVQFTIINGGGTSFSYDERNLVSSL	1881
Qy	1538	-----DAIINNLLVYVNTTGGGEYQKKYGGAPLDKLOKL-----	1571
Db	1882	HIGDKNGGDILTESYEYDA---NGNRRTINSSASGKVQYEG-----KLNQLVKRETHEDGT	1934
Qy	1572	-----YPEIFTKQVSTGVAIDPDSQKITESAKYFNCT-NILHR-----GS	1611
Db	1935	VIEITYDGFGRKVIYTT-----TKDGSKIVNASFINMNLTKVNDESISYDRKNGN	1985
Qy	1612	GYVLKADGGQY-----NLGTTTKQFLPQLTGTGKQGNFVKGNNGNY--YFYDLAG	1663
Db	1986	---RTSDGKFTYTWDADENLTAVTK-----KGEDKPFATYKYDEKG	2023
Qy	1664	NMVKNTIFIEDSVGNWYFFDDQDGKMYENKHFVDV-----SYGKGTGYFFFLKNGVSFRG	1716
Db	2024	NRIOKT-VNGKVTNIFY---DGDLSNLVLETADADNNVTKSYTYGDSGQL-----	2068
Qy	1717	GLVQTDNGTYTFDNKGYKVRNQTINA-GAMIYTLBENGKLIKASYNDSA-EYPTSTDYVK	1774
Db	2069	-LSYTEGNGKKEVHY-----NAHGDIIAISDSTGKTV-AKYQYDANGNPTKTEASD	2117
Qy	1775	MLDONK	1780
Db	2118	EVADNR	2123
RESULT 8			
APU_THETU			
ID	APU_THETU	STANDARD;	PRT; 1861 AA.
AC	P3836;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	AmiOpullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase		
DE	type II) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan		
DE	glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan		
DE	glucanohydrolase) (Alpha-dextrin endo-1,5-alpha-glucosidase)].		
GN	AmiB.		
OS	Thermoanaerobacter thermosulfurogenes (Clostridium		
OS	thermosulfurogenes).		
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;		
OC	Thermoanaerobacteriaceae; Thermoanaerobacterium.		
OX	NCBI_TaxID=33950;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-DSM 3896 / EMI;		
RX	MEDLINE=9425298; PubMed=8195085;		
RA	Matuschek M., Burchhardt G., Sahm K., Bahl H.;		
RT	"Pullulanase of Thermoanaerobacterium thermosulfurigenes EMI		
RT	(Clostridium thermosulfurogenes): molecular analysis of the gene,		
RT	composite structure of the enzyme, and a common model for its		
RT	attachment to the cell surface";		
RL	J. Bacteriol. 176:3295-3302(1994).		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic		
CC	linkages in oligosaccharides and polysaccharides.		
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic		
CC	linkages in pullulan and in amylopectin and glycogen, and the		
CC	alpha- and beta-limit dextrins of amylopectin and glycogen.		
CC	-!- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN		
CC	S-LAYER ANCHOR.		
CC	-!- PTM: GLYCOSYLATED.		
CC	-!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO		
CC	KNOWN AS THE ALPHA-AMYLASE FAMILY.		
CC	-!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.		
CC	-!- SIMILARITY: CONTAINS 3 S-LAYER HOMOLGY (SLH) DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; M57692; AAB00841.1; .		
DR	HSSP; Q08751; 1BVZ.		
DR	InterPro; IPR000461; Alpha.amylase.		
DR	InterPro; IPR004185; Alpha.amylase_N.		
DR	InterPro; IPR003961; FN.III.		
DR	InterPro; IPR004193; Isoamylase_N.		
DR	InterPro; IPR001119; SLH.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00128; alpha-amylase; 1.		
DR	Pfam; PF00395; SLH; 3.		
DR	Pfam; PF02806; alpha-amylase_C; 1.		
DR	Pfam; PF02903; alpha-amylase_N; 1.		

Pfam: PF02922: isomylase\_N; 1.  
 DR SMART; SW00060: FN3; 1.  
 DR PROSITE; PS01072; SLH\_DOMAIN; 3.  
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;  
 MW Multifunctional enzyme; Glycoprotein.  
 FT CHAIN 1 35 POTENTIAL.  
 FT TAG 36 1861 AMYLOPULLULANASE.  
 FT DOMAIN 928 1018 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 1157 1248 FIBRONECTIN TYPE-III 2.  
 FT ACT\_SITE 628 628 BY SIMILARITY.  
 FT ACT\_SITE 657 657 BY SIMILARITY.  
 FT ACT\_SITE 734 734 BY SIMILARITY.  
 FT DOMAIN 1681 1739 SLH 1.  
 FT DOMAIN 1740 1803 SLH 2.  
 FT DOMAIN 1804 1861 SLH 3.  
 FT CONFLICT 1734 1734 D -> E (IN REF. 1; AAB00841).  
 SQ SEQUENCE 1861 AA; 206104 MW; 06C23070E453B574 CRC64;  
 Query Match 3.1%; Score 287.5; DB 1; Length 1861;  
 Best local similarity 18.5%; Pred. No. 9e-06;  
 Matches 324; Conservative 236; Mismatches 561; Indels 635; Gaps 91;  
 33 GGVAHQDQVQQAQTQDQSTVNNDTKUVALDNTD-----QSA 73  
 358 GGVGKA-----TDTVKDFELIV-YDKNFDTPDMKGVAMYQIFPDRFNGD 403  
 74 QTDDKKQVYNTNOSKTDSTAD-----KNSTSTP----- 104  
 404 TSNDHAKTLRGNDPIEFHNWNLDPNPNNAGTPGYTGDIWSNDFGDLKGIDDKLDY 463  
 105 -----VSLPSNTEKQAKYNEODKNGYGNIDTAYPSN-----NQLHVSQWNAINAS 152  
 464 LKGLGVSVIYNLPFEPSNH-KYDTADYKIDEMFGTTQDFELKMSDAHAKGK----- 517  
 153 QGTNSRQLIVRDTNNELGRTDVTNNVARPDVKVNHVYNADNSG-----FDVNVNI 205  
 518 -----IILDGVFNHTSDSIYFNRYGKYPGLGAYQANKEGNSQSLSPYGDWYTIISDG 569  
 206 DFKMKDYRDSIEIVRSYSGNGSKVDWKSQIPITDKNYA--YLD-----TFE 251  
 570 TYECWNGY-DSLPIVKSUNGSEYNTSWANFIINDENAIKYLWNPQGNLNDGADGWRLD 628  
 252 VKNGELH-----ATGW-----NATNSAINYRHHFVIL----- 278  
 629 VENEVAHDFWTHFRNAINTVKFEAPMIAENNGDASLDLLGDSFNSVMNYQFRNDIIDFLI 688  
 279 ---FDQTNGKEVARQEVREGSRPDVAKYPOVYGAANSFNVTFNISDLDTYHOYQVLS 335  
 689 GQSFDDGGGQHNPIIDAAKLDQRLMSIYERYPL-----PAPYSTNNLLGSHDTMRILTVF 742  
 336 RYSNSDNGEGDNTYWFENPQSIAPANGSNQYLDSDISKNGEYTVTG-----W 384  
 743 GYNSADPNENSDAKQLAEQKLKATILQMGYPGMADIYGDAGVSGGKPPDDRRTFPW 802  
 385 NATD-----LSELQNNHYVILFDQTACKQVASAKADLISRPDVAKAYPTVKYATNS 435  
 803 GNEDTTLQDFFKNISIRNNQVL---KTGDELTLYAQNDV-----ID 489  
 436 GFKYTFKVNLLQPGHQSVSRFSADENGNGDKRHTDYFSPVI--LNOTASN----- 840  
 841 -----YAIGRII-----NGKDAFGTSYPSDAIAIVAINRSKSKQIAID 879  
 490 TITWTSNGLHIAGWASDNSINETTPYAILNNGKEVTRQKMSLTARPDVAAYPSLYNS 549  
 880 TTKFLRDG-----VTFKDLINNVSYSI--SNGQIVI-----DV-----P 912  
 550 AVSGFDITIKLTNDQYQALNQQLQVLLRFSKAADGN---PSGDNVTVDQFSKNYATTTGGN 606  
 913 AMSG-----VMLISDDGQDLT-----APQAPSNVVTSGNGKVDLSWLQSDGATGYN 959  
 607 FDYVKVNGQVFFSGWHATNOSDKDSQWIIIVLNG-KEYKRQLVNDTKEGAAGFNNDV 665  
 960 IYRSSVEGLYEKIASNVT-ETTEDAN-----VTNGLKYVYVAISAIDELGNESGISNDV 1014

QY 666 -YKVNPAIENSSKSGFQGIITLPTVTKNENYQLVHRFSNDYKVTGEGNVYDFWSELMPYKD 724  
 DB 1015 AYPAYPI-----GWVGNLT--QVSDNHIIGVDPKTFEDI-----YAEVAD--GUTN 1056  
 QY 725 SFQKGNGLPKQFGLQTINGQYYIDPTTQGPKNFLLOSNNWYFDSDTG-VGTNALEL 783  
 DB 1057 STGQGFNMIAQLGKYKVSQTVI--DSVTG-----SVYNSVGVDDSGFTWNA--- 1102  
 QY 784 QPAKGTVSSNEQYRNGNAAYSYDDKSIENVNGYLTADTWYRPKQILKDGTTWDSKETDM 843  
 DB 1103 QYV-GDIGNDQYK--ASFTPD-----KIGQW---EYLMR-----FSDNQGD- 1139  
 QY 844 RPLMWNPNTLTQAYYLYNMKQHGNNLPSALPFFNADAPAEALNHYSEIYQONTIE-KRI 902  
 DB 1140 -----WITSTLSF---YVVPSSDILLIKPTAPYLN-----QPGTESSRV 1174  
 QY 903 SETGNTDMLRTLMHDFVTNNPMWKNKDSNVNFSGIQFGGFLKYENSDDLTPVANSDYRL 962  
 DB 1175 SLT-----WNPSTDNV-----GIYDYE-----IYRSGGTF 1200  
 QY 963 GRMPINIKDQYRGQEFLLANDIDNSPVYQAEQLNWLXYLLNFGTITANNDQANFDSYR 1022  
 DB 1201 NKI-ATVSNEV-----NYIDTS--VINGVTYNYKVAVD--LSFNRTESNVVTIK 1246  
 QY 1023 VD-----APNIDADLMNIAQDFNAAAYGMDSDAVSNKHINILE----- 1061  
 DB 1247 PDVVPKIVFNVTVPD-YTPDAVNLAGTFPNATWDPFSAQQMTKIDNNTYISITLDEGTQ 1305  
 QY 1062 -----DWNHAD-PEYENKI-----GNPQLTMDDTIKNSLHGLSDATNRWG 1101  
 DB 1306 IEYKARGSWDKVEKDEYGNFASNKVITVNOGNEMINDTV-----YRW- 1352  
 QY 1102 LDAIVHQSADRENSTENVVTPNYSFVRAHNNNSODQIQONAIRDVTGDKYHTTFEDSQ 1161  
 DB 1353 -----RD-----IP--IFIYSPSNMTVDSNISTMEVKGNFY----- 1382  
 QY 1162 KGI-----DAYIQDNSTVKRYNLYNIPASVAILLTNKTDTI---PRVYDGLYTDGGQY 1212  
 DB 1383 KGAKVTINGDSFQDANGVETK---DVSINYGCV---NKIKIHVEP-----NDGSYV 1427  
 QY 1213 MEHQTRYDDELTLNLLKSRVYVAGGOSMOTMSVGGNNNLTSVRYGKGAMTADTGTDET 1272  
 DB 1428 GNDQGR-----ITELTKD-IEIDVIRGNSSGSGTGNNNTSTS-----GSNSSTSGST 1476  
 QY 1273 RTQIGIGVYVSTNPKL-LGV-----NDRKVLHMGAAHKN 1305  
 DB 1477 GSTISITSNTSNTSNTTIGVITKGNVITLILDAGKADLVNSKDKKVVFDITIGEG 1536  
 QY 1306 QYRAAVLTTTIDGVINYTSDDQAPVAMTDENGDLXLS-----SHNLVYNGKEEADTAVQY 1361  
 DB 1537 QQ---KVVQISKDILTSAANGKDIVIKSDNASIALTKDALNQNLQNG---VNVSIKDN 1590  
 QY 1362 ANPDVSGYLAVWYV-----GASDNQDARTAPSTENKSGNSAVRTNAAFDSNVIFEAFN 1416  
 DB 1591 GKPNVINYSLSNVVDITISIGISGNVTLAKPVEVTLN-----ISKANDPRKV----AV 1639  
 QY 1417 FVYTPKES-----ERANVRIAQNADFFASLGFTSFEMAPQYNSKD----- 1458  
 DB 1640 YYNPITTNQWEYVGGKVDASSGTTFNATHFSQ--YAAPEYDKTFENDIKDNWAKDVIEVL 1697  
 QY 1459 --RTFLDSTIDNGYAFTRDYLQMSFNKYGT-----DE-----DLR- 1493  
 DB 1698 ASRHIVEGMDTQY-----EPNKTVTRAETAMILRLNLKIDETYSGEFSDVKS 1746  
 QY 1494 -----NAIOALHAKGL 1504  
 DB 1747 GDWYANAIEAAYKAGI 1762

RESULT 9  
 ASAL\_ENTFA  
 ID ASAL\_ENTFA STANDARD; PRT; 1296 AA.

AC PI7953;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aggregation substance precursor.  
 GN ABA1.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OG Plasmid PAD1.  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-51.  
 RC STRAIN=DS16;  
 RX MEDLINE=91014689; PubMed=2120541;  
 RA Galli D., Lottspeich F., Wirth R.;  
 RT "Sequence analysis of Enterococcus faecalis aggregation substance  
 RL encoded by the sex pheromone plasmid PAD1.";  
 RL Mol. Microbiol. 4:895-904(1990).  
 CC -!- FUNCTION: AGGREGATION SUBSTANCE ALLOWS DONOR AND RECIPIENT STRAINS  
 CC TO FORM TIGHT AGGREGATES WHICH ALLOW THE NON-MOTILE BACTERIA TO  
 CC MAINTAIN PHYSICAL CONTACT OVER A PERIOD OF TIME SUFFICIENT TO  
 CC PERMIT CONJUGATIVE TRANSFER OF THE SEX PHEROMONE PLASMID FROM  
 CC DONOR TO RECIPIENT STRAINS.  
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 CC an amide bond (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X17214; CA35083.1; -;  
 DR PIR; S10223; HMSOIF.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR TIGRFAMS; TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 KW Cell wall; Peptidoglycan-anchor; Signal; Plasmid.  
 FT SIGNAL 1 43  
 FT CHAIN 44 1264 AGGREGATION SUBSTANCE.  
 FT PROPEP 1265 1296 REMOVED BY SORTASE (POTENTIAL).  
 FT SITE 1261 1265 LPXTG SORTING SIGNAL (POTENTIAL).  
 FT MOD\_RES 1264 1264 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
 SQ SEQUENCE 1296 AA; 142285 MW; 52123A13AD23E5B CRC64;  
 Query Match 3.0%; Score 286.5; DB 1; Length 1296;  
 Best Local Similarity 20.2%; Pred. No. 6.1e-06;  
 Matches 292; Conservative 170; Mismatches 493; Indels 491; Gaps 70;  
 QY 2 EIKKHFLYKSGKOWTAATVAATVASTALLYGGVAHADQVQAS-TTODOTSTVNNDD 60  
 DB 6 EVKRFMYAKKHVVAPILFTGV-----LGVVGLATDDVQAAELDTQGGTTVQPD-- 58  
 QY 61 KTVALDINTDOAQTDDKQVSNNTNOSKTDSTADKNSTSTPVSPLSNNTKQAKNY 120  
 DB 59 -----NPD-----PQVSTTPKTAETEAQKDTTSQTKV-----EEVASEK 97  
 QY 121 NEODKNGYGNIDTAYFNNLHVSNGWATNASOGT-----NSRQITIVRDIITNNELGR- 173  
 DB 98 NGAEQSATPNDT-----TNAQPTVGAEKSAQEQPVSPETNEPLGQP 142  
 QY 174 -----TDVTNNVARDPVKNHVNYNADNSGFDVNNIDFSKMKDYRDSIEIV 220  
 DB 143 TEVAPAEANEANKSTSPKEFETPDVDKA-----VDEAKKDPNITV-----VEKP 186  
 QY 221 SRYSGNKSVDWNSOPTTFKNNYAYLDTEVKNGELHATG-----WNATNSAIN----- 270  
 DB 187 AEDLGNVSSKDLAAKEKEVDOLQEQAKKTAQQAALAKNEKIAKENAIAAKNAEKE 246  
 QY 271 -----YNHH-----FVILEDQ-----TNGKEVARQ-----EVREQSRPDVA 302

DB 247 RYEKEVAEYNKHKNGYVAKPVNKTILFDREATKNSKVSKAAEYIDAKKLTDKHKDK 306  
 QY 303 KVPQVGAANGFNWTFENISDLDTYHOYVLSRYSNSDNGEDNVYTFWNPQSIAPANQ 362  
 DB 307 KLLISMLSVDSGLTT---KDSKKAHFY-----NNGAGGTLYV----- 342  
 QY 363 SNOGYLDSFDISKNGEVTWGNATDLSLONNHVILFDQTAGKOVASAKADLISRPDV 422  
 DB 343 -----LHKNOPVITIGNL-----NASYL-----GKTIASAEFY-----T 373  
 QY 423 AKAYPTVKATNSGF-----KVTFKVN-----NLQPGHOYSVVSRSFADENGNGDKRHT 472  
 DB 374 VKATPDSKGRNLAFLDDDPVATIVGIVINIDPRTKKAGAEIEMLVRFEGDGEILPTKEN 433  
 QY 473 DWYFSPVILNQTSNI-----DI-----ITMSNGLHIA 501  
 DB 434 PFVSGASLNSGENITFEYVKVGNNTDTHHEINGSKVARHGNKVYSKTDIDVGTNGISIS 493  
 QY 502 GWMASDNS--INET---TPYAILNNGKEVTRQKMSLTARPDVAAYVPSLYNSAVSGFDT 556  
 DB 494 DWEAVQGEYIGATVISTPNRIKFTFGNEIYN-----NPGYDG-----NSMWFAPNT 540  
 QY 557 TIKLTN-----DOYQALNGOLQVLLRFSKRAADNPS-GDNTVTDQ 595  
 DB 541 DLKAKSITPYQEKGRPKQPEKATIEFNRYKA--NVVPLVPNKVTDGQKINIDLNVKRG 598  
 QY 596 FSKNYATTGGNPDYKVNNGNOVEFSG-----WHATNOSNOKDSQWI 636  
 DB 599 DLSQYIVTGDITELAKVDPKVTYKOGIRDTDAEKVTIDLSKVYVQADASINELDKLAV 658  
 QY 637 IYLVNGKEVKROLVNDTVEGAAGFNRNDYKVNPAIENSSM---SGFQGITLPTVYK-- 692  
 DB 659 AAAINSGRKAK---DVTASYDLHLDONTVATMMKTNADDSVVLAMGYKYLLVLPVVKNV 714  
 QY 693 ----ENVQLVHRFSNDVKTGEGNYVDFWSELMPVKDSFKQNGPLKQFGLQTINQOQYI 748  
 DB 715 EGDENTAV--QLTNDGETVTVTVINHVPNSPKDVKADKNGTVGSVSLHD----- 764  
 QY 749 DPTTQPKRNFLLQSGNNWIFDSDTGVGTNALELOFAKGVSSNEQYRN--GNAYSVD-- 806  
 DB 765 -----KDIPLQTK---IYE-----VKSSERPANTGGITEWGM 795  
 QY 807 ----DKSTENVNGYLTADTWYRPK---QLKDGTTWTDSDKETDMRPILMWVWPNLTQAY 859  
 DB 796 NDVLTTHDRFTGKHATNTYDLKVGDKTLKAGT-----ISAY 834  
 QY 860 YLNYKQKGNLLPSALPFNADADPAELNHYSEIVQ-----NIEKRISE 904  
 DB 835 ILLENKDNKD-----LTFTMNOALLAALNEGSNKVKGQAWSVYLEVERIKTGDVENTOTE 889  
 QY 905 TGNTDWTFLTMHDFVTNNP-----MWNKDSENVNFSGIFQGFGLKYSN-DLTPYA 955  
 DB 890 NYNKELVRS--NTVVTHTPDDPKPKAVHKKGEDINGKVA-RGDVLSYEMTWDLKGYD 946  
 QY 956 N-----SDYRLLGRMPI-----NIKQTYRQGEFLLANDIDSNPNVQAE 995  
 DB 947 KDFAEFTVDLATGVSFFDDYDEYKVTPIKDLRVRDS--KGE-----DITNQFTISWDD 998  
 QY 996 QLNWLYLLNFCGTTITANNDOA-----NFDSEVRVDAPDNIDALMNIADQYENAA-----YG 1046  
 DB 999 AKGTV-----TISAKDPQAFILAHGGOELRVTLPTKVA---NVSGDVNLABQNTFG 1048  
 QY 1047 --MDSDAVSNKHINILEDNNHADPEYFNKIGNPQLTMDDTIK-----NSLNHGL 1093  
 DB 1049 ORIKNTVYN-HIPKV---NPKDWIKVKGDKQSONGATIKLGEFFYEFTSSDIPAEY 1103  
 QY 1094 SDATNRWGLDAIVHOSLADR-----ENNSTENVLPNPSFYRAHDN--NSODOIQNAIRD 1146  
 DB 1104 AGIVEEW-----SISDKLDVKHDKFSGQMSVFANSTFVLADGTVKNKGDDISKLP-- 1153  
 QY 1147 VTGKYDHTTFEDEQKI-----DAYIQDNQSTVKYKYNLXNIPASAYAILTNTKDTIPRV 1200

Db 1154 -----TMTFE-----QGVVKITASQAFDAMNLKEN-----NVAHSMKAFI-----GVERI 1195  
QY 1201 YYGDLV 1206  
Db 1196 AAGDVI 1201  
RESULT 10  
YD\_BA\_ECOLI STANDARD; PRT; 2003 AA.  
AC P3366; P76087; P76085; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ydBA.  
GN YD\_BA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,  
RA Kasai H., Kishimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1995).  
[3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190333; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS6 (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2B/IS30C ELEMENT  
BETWEEN AMINO ACIDS 339 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000237; AAC74483.1; ALT\_SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT\_SEQ.  
DR EMBL; D90778; BAA15009.1; ALT\_SEQ.  
DR EMBL; D90778; BAA18880.1; ALT\_SEQ.  
DR EMBL; D90779; BAA18881.1; ALT\_SEQ.  
DR EMBL; X62680.; NOT\_ANNOTATED\_CDS.  
DR EcoGene; Egl1307; ydBA.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;  
Query Match 3.0%; Score 285.5; DB 1; Length 2003;  
Best Local Similarity 19.2%; Pred. No. 1.2e-05;  
Matches 380; Conservative 225; Mismatches 718; Indels 655; Gaps 96;  
QY 65 LDTNTDQSAQTDDKKQVSVNTNQSKTDDTSTA-----DKNSTSTPVSVLPSSN- 112  
Db 86 VELNDDDDHHRNNSPLPTPDDESDDTPVPTPGGDBIIPDDPDDTPTPPKPVSEFND 145  
QY 113 -----TEK-----QAKYNEQDKG-----NYGNIDTAYFSNNQLHVSQWN 147  
Db 146 VILDKTEKTLTIRDSVFTYTNADGTISLQDSNGRKATINLWQIDEA---NNTVALEGVS 202  
QY 148 ATNASQ---GTNSRQIIVRDITTNBELGTDVTNNVARDVKNVHVNADSGEDVAVN 204  
Db 203 ADGATKQYNHNGELVITGDNATVNNNGKTYDVGK-----DSTGTEINGN 247  
QY 205 IDFSKMKYRD-SIEIVSRYSNGKSVDMWSOPITFDKNNYAYLDTFFVKNGELHATQWN 263  
Db 248 ----NGKVIQDGDLDV----SGGGHGIDITGDSATVDNKG-----TMTVDPESMGID 294  
QY 264 ATNSAIFYNHFFVI-----LFDOTNGKEVARQEVREGQRPDVAKVYPQVVGAAANGFNT 319  
Db 295 CDKAIYNNEGESTIINGCTGTGTOINGDDATAN--NNGKTTVD-----GKDSGTGTEIN 343  
QY 320 FN-----ISLDYT---HQYQVLSKYSNSDNGEGDNTVYWFNPQSIAPANSQGYLDS 370  
Db 344 GNGKVIQDGDLDVSGGGHGIDITGDSATVDNKGTMVT---DPESIGIQVGDQDAV--- 397  
QY 371 EDISKNGEVTVT-GNNATDLS---ELQNNHYVILFDQTAGKQVAKASAKADLISRPDAKAY 426  
Db 398 --VNNEGESAITNGTGTGTOINGDDATANN-----GK----- 427  
QY 427 PTVKTAINGFKVTFKVNKLQPHQYSVVRSFADENGNDKRDHTYWFSPVILNQRTAS 486  
Db 428 TTVDGKDSGTGTEA-----GNGKVIQDGDLDVSGGGHGIDITGD-----SATVD 472  
QY 487 NIDTITMT---SNGLHAGMASDASINETTPYAILKNGKEVTRQKMSLTARPDVAAYV 543  
Db 473 NGKTMVTDPESIGIQDGDQAIYNNEGEST-----ITNGGT----- 509  
QY 544 PSLYNSAVSGFDTTKLTNDYQALNGQLVLLRFSKAADGNPSGDNTVDTQFSKNYATT 603  
Db 510 ----GTQINGNDAT-----ANNSGKTTVDGKDSGTG-TKI 538  
QY 604 GGNFDYKVKVNGQVFEFSGHATNQSD-----KDSQWIIVLVNGKEVARQLVN 651  
Db 539 AGNIGIVNLGSLTWTGGAHGVENICDNGTVNNKGDIVVSDTSGISGLVING----- 589  
QY 652 DKKEGAAGFNRRNDVYKVNPA-----IENSSMSGFOGIITL-----PVTYKN 692  
Db 590 ---EGATVSTNGDVVSVNEATGFTSTNSGKVSAGSAGVQDFSTGVDLNGNNSVTLAA 646  
QY 693 ENVLVHRESNVK-TGEGNYVDFWSELMPVKDSFKQKNGPLKQFGLQTINQYQYIDPT 751  
Db 647 KDLKVVQKATGIVSGDANTVNTGNVLVDKDK-----TADNAAEYFFDPS 693  
QY 752 TG---QPRKNELLQSGNNKIYFSDT-----GVG-----TNALE 782  
Db 694 VGIVNYGSDNNVTLDGLKLTVDSDSEVTSRQSNLFDGSAREKTSGLVIVGDTGNVNGGLE 753  
QY 783 LQFAKGTSSNQYRNGNAAYSDDKSIENVNG----YLTADTWYRPKQILKDGTTWTD 838  
Db 754 LIGEKNALADGSOVTSRLTGYSY--TSVIVSGESSVVLNGDT----- 794  
QY 839 KETDMRPILVWVWPNLTLCAYVINY----MKOHGNLLSALPFFFNADADPAELNHYSIV 894  
Db 795 -----TISGEFFLGFAGVIRVDKALL-----EIGSGATLT 825  
QY 895 QONIEKRISSETGNTDLWRLTMHDFVTNNPMWKNKDSENVNFSGIQ-FQGG-----FL 944  
FT CONFLICT 489 489 I -> V (IN REF. 2)









RESULT 12  
OMP\_RICJA STANDARD; PRT; 1656 AA.  
AC O06653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)  
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein  
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptidel.  
GN OMPB.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH;  
RA Uchiyama T.;  
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
RT japonica.";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
CC SIMILARITY)  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC use European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AB003681; BAA20138.1; -  
DR InterPro; IPR003858; rOmpA\_rOmpB.  
DR Pfam; PF02708; rOmpA\_rOmpB; 1.  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.  
FT DOMAIN 528 533 POLY-GLY.  
FT DOMAIN 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;  
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;  
  
Query Match 2.9%; Score 274.5; DB 1; Length 1656;  
Best Local Similarity 20.2%; Pred. No. 3.1e-05;  
Matches 413; Conservative 218; Mismatches 732; Indels 681; Gaps 107;  
  
QY 1 MEIKKHF--KLYKSKQWVTAATVAVSTALLYGGVAHQDQVQQAQTQDQSTVNN- 57  
DB 1 MAQKPNFLKLIISAG--LVTAATATIVASFAGSAMGAA-----IQQRRTNGVATTVDGV 53  
QY 58 DTDKTVLDTNTDQAQTTDKKVVSTNQSKTDDTSTADKN-----STSTPVSVLPS--- 110  
DB 54 GFDQPVAL-----ANVAVPNAVITANANGINLNTFAGSFNGLFLSNANNAVTVSEDT 108  
QY 111 -----NTEKQAKNVN-EQDKG-----NYGNIDTAFSNQHLVSGWN-----ATN- 150  
DB 109 TLGFINNAANNANRFLDLAGKLTITIGQGITNVQSAATHNAQNAIVAKFNFGAAIANNND 168  
QY 151 -----ASQGTNSRQIIVRD--ITTNELGRTDVTN--NVAREPV 185  
DB 169 LSLGLTIDFGAASTLVPDLANPTQKAPLILADNALIVGANCATLNVNTNGFTQVSKSF 228  
QY 186 KNVHVNVNADSGFDVNVN-----IDFSKMDYRDSIETIVSRYSNGKSV 230  
DB 186 KNVHVNVNADSGFDVNVN-----IDFSKMDYRDSIETIVSRYSNGKSV 230

DB 229 ATVKAINIGDGGQGFMTNATNANALNLQAGGTTINFNG--TDCTGRLLVLSK---NGAAT 284  
QY 231 DW-WSQPIITFDKNYAYLDTFVKNGELHATGNA-----TNSAINYHHFVILFDQNG 284  
DB 285 DFNVTSGLGCLNKGIIELNVAI--NGQLIANAGPANAVIGTNGAGRAAGFVVSVD--NG 341  
QY 285 KEVARQEVREGSRPDVAKVYPO--VYGAANGFNVTF--NISLDYTHQYQVLSRYSNSD 341  
DB 342 KAA-----TIDGQ-----VYAKDMVIQSANANGQVNFRIHVDVGI----- 376  
QY 342 NGEQDNVTVWFENQSTAPANO--SNOGYLD-----SPDISKNGEVTGTGNATLSELQNNH 396  
DB 377 -----DGTAFKTAASIVAITQNSNFTGFGNLAQVTPDVTMTLTGNTGDANNGNTA 432  
QY 397 YVLFQDTAGKQVYASAKADL-----ISRPDVAKAYPTVKTATNSG--FK- 438  
DB 433 GVITF--AANGTILASASADANAVTNNTAIEASGVVQLSCTHTAELRLNAGSVFKL 490  
QY 439 -----VTEKVN-----LQGHQVSVVSRFSADENGNGDKRHTDYFSPVILNQTA 486  
DB 491 ADGTVNGKVNQTVLVGGVLAGAITLQGSATITDIGNGG-----GAALQSITILA 542  
QY 487 NIDTITMTSNGLIAGWASDINSINETTPYAILNNGK--EVTRQKMSLPARPDAVAVPS 545  
DB 543 NDATKTLTLGGANIIS--ANGGTIN-----FOANGGTIKLTSTQNNIVVDCDLAIA--- 591  
QY 546 LYN SAVSGFDTTITKLNDQYQALNGOLOVLLRFSKAADGNPSGDNTVTDF---SKNYAT 602  
DB 592 ---:DTGTGVVDDASLTNAQTLTISGTIGII-----GANNTLQGFNIGSSKTTL 637  
QY 603 TGSNF---DYVKVNGQVFEFGWHAT---NQSNKDSQWIIVLVNGKEYKRLVNDTKEG 656  
DB 638 NGGNVAINELVINGNSVQFA--HNTYLTIRTTNAAQ-----GKIIFNPVNNNTL 688  
QY 657 AAGFNNDYKVPATE--NSSMSGFQGIITLPVYKNEVNVQLVHRPSNDVKTEGNYVDF 715  
DB 689 AASTNLGSA--ANPLAEINFGSKGARADVLNV---GEGVNL---YATNITTTDANVGSF 740  
QY 716 WSELMPVKDSFQKNGPLKQFGIQTINGQQYITDPTGTQPRKNFLQSGNNIYFSDSTG 775  
DB 741 -----VFNAGG--KNIVSGTVGGQ-----GNKENTVALDNG 770  
QY 776 VGTNALELQFAKGTSSNEQYRNGNAAYSDDKSIENVNGYLTADTWYRKILKDGTTW 835  
DB 771 TTVKFL-----GNATF--NGNTTIAAN---STLQISGNTADF-----TASADGTGI 812  
QY 836 TDSKETDMRPILMVWVWPNLT--QAYVLYNMKQ-----HGNLLPSALPFPFNADAPLNLH 889  
DB 813 VEFVNTG--PINV-----TLNKQAVPVNALKQITVSGPGNVV----- 847  
QY 890 YSEIVQONIEKRISGTGNTDMLTLMHDFVTNNPMWNKDSENVNFSGIOFQGGFLKYENS 949  
DB 848 -----VNEIGNAGNVHGAMTDITIA-----FENSSGLAVFLPSPGPFNDA 887  
QY 950 DLTPYANSDYRLGLRWPINK---DOTYRG-----QEFLLA 982  
DB 888 GNT-----IPLTIKSTVGNETAEGFSVPVIVSGVDSVTADQGVTDQNNIYV 935  
QY 983 NIDNSNPVYVQBLNWLXYLLNFGTITANNQANPDSVRVADPNIDADLNMIAQDYN 1042  
DB 936 LGLGSDNGII---VNATTLIYAGIGTINNQGVTLSGGVPNTPGTV----- 978  
QY 1043 AAYGMSD--DAVSNKHINILEDWADPEYFNKIGNPQLTMDDTIKNSLNHGLSDATNW 1100  
DB 979 --YGLGTGIGASKFKQVTFITD-----YNNLGNIIAT-----NTTINDGVTVTIG-- 1021  
QY 1101 GLDA-----IVHQSADREN-----NSTENVVIPNYSFVRAHDNNSQDOI 1140  
DB 1022 GIAAGTAGTDFDQKITLGSVNCNANVRADGIFSNSTNIV-----TTKANNGVTVL 1075  
QY 1141 QNAIRDVTKDYHT-----FTFEDEQKIDAYIQDNQSTVKYKNLYNPASVAILLTK 1194  
DB 1076 GNAFVGNIG--DSDTPVASVRFTGSSNNGAGLKGNIIYSQVDFGTYNL--GIVNSNVL-- 1129

```
QY 1195 DTIPRVYGDLYTDGQYMEHOTRYDYTLNLLSRVYKVVAGGSMOTMSVGGNNILTS 1254
DQ 1130 -----GGSTTAINGK-IDLLNTLT-----PAGG-----TSTWGNNTSTETT 1165
QY 1255 VRYGKG-----AMTATDGTDETRTOGIGVVVSTPNLKLGVNDKVLHMGAAH 1303
DQ 1166 LTLANGNIGHIVIAEGAQVNAATTGTTT-----INVQDANANFSGTOTYLLIOGAR 1218
QY 1304 KNOOYRAAVLTMT--DGVINY--TSDQAPVAMTDENGDLYLSSHNLVYNGKEEADTA 1357
DQ 1219 FNGTLGGPNEFTVTGSNRFVYGLLRAANQDVIITRN-----NAENIVTN----- 1263
QY 1358 VQGANPDVSYLAWTPVPGASDQDARTAPSTKNSGNSAYRNAAFDSNVIFEAPS-- 1415
DQ 1264 -----DIIN-----SFFGG-----APGVGNVTFVFNATNAYNNLLAKNSAD 1303
QY 1416 --NFVYPTKESERA--NVRIAQNADEFASLGFTSFEMAPQYNSKORTFLDSTIDNGYA 1471
DQ 1304 SANFVGTVTDTSAITNAQDLVAKDIAQQLGNRLGAL----- 1341
QY 1472 FTDYDILGMSPNKYGTDEDLRNAIOALHKAGLOVMDWTPDQIYNLPGKEVATVTRVDD 1531
DQ 1342 ---RY---LGTPEMVGSP---AGALPAVAAGDEA---VDNVAIGIWAQFFYTDHQSK 1388
QY 1532 RGN-----VNMKDAIINNLYVNTIG-----GGEYQKRYGGAFLDKLO--- 1569
DQ 1389 KGLAGYKAKTGTGIGLDTLANNLMIGAAGITKTDIKHQDYKKG-----DKTDVNG 1442
QY 1570 ---KLY-PEITTKQVSTGVAIDSQKITEWSAKYFNGTNILHRGSGYVLKADGGQYINL 1625
DQ 1443 FSELYGAQQEVENFFAAGSFAISLNVQVKNKSQRFFDAN----- 1482
QY 1626 GTTTKQFLPIQLTGEKKQGNFVKGDNGNYF-----YDLAGNMVKNFTFEDSDVGNV 1678
DQ 1483 GNMKSQ-----IAAGNVDMNFTGSLTVGYDY--NANQGVLTVPMAGLS 1524
QY 1679 YFFDODGKMNKHVFDVDSYGEKGTFFFLAN-----GVSFRCGLVQVOTNGT 1725
DQ 1525 YLKSSD-----ENYKETGTTVANKQVNSKFSRDTDLIVGAKVAGGTMNITDFA 1572
QY 1726 YFFONY-----GKVMRNQINAGAMIYTLDENGKLIRASYN-----SDA---EYP 1767
DQ 1573 YYPEVHAFVHKVGRLSKTSVLDDGVQVTPCISOPDRKTSYNLGLSASIRSDAKMEYG 1632
QY 1768 TSTD 1771
DQ 1633 IGYD 1636

RESULT 13
OMP_B_RICCN STANDARD; PRT; 1655 AA.
AC Q9KKA3; Q9KX98; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)
DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR RC1085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
```

```
RT Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN [3]
RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008659; AAL03623.1; -
DR EMBL; AF123721; AAF34124.1; -
DR EMBL; AF123726; AAF34129.1; -
DR EMBL; AF149110; AAD39533.1; -
DR InterPro; IPR003858; rOMPb_rOMPb.
DR Pfam; PF02708; rOMPb_rOMPb; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.
FT VARIANT 61 1655 32 KDA BETA PEPTIDE.
FT VARIANT 75 75 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 KD -> GH (IN REF. 3).
FT CONFLICT 776 776 F -> S (IN REF. 3).
FT CONFLICT 1159 1159 E -> D (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match 2.9%; Score 269; DB 1; Length 1655;
Best Local Similarity 20.5%; Pred. No. 5.5e-05;
Matches 403; Conservative 216; Mismatches 700; Indels 650; Gaps 106;

QY 1 MEIKKHF--KLYKSGQWVTAATVAVSTALLYGVHADQVQQAATQ----- 49
DQ 1 MAQKPNFLKLLISAG--LVTAATATIVASPAAGSAGAA-----IQNRTTNAVATVDGV 53
QY 50 --DOTSVNN-----DQDKIVALDT-----NTDQSAQITDKK----- 79
DQ 54 GFQDQATVAVPLNAVITAGVNGKITLTPAGSFGNGLFANNLDVTVREDTLGLFI 113
QY 80 -QVYSNTN-----QSKTDDTSTADKNSTSTPVSVLPSSNTEKQAKN--- 119
```

Db 114 TNYVNAHFNLMLNAGKTLITGQGITNVQAAATKN-----ANNVVAQVNNGAA 163  
QY 120 YNEQDKGNYGNIDTAYFSNNOLHYSGMNATNASQGTNSRQIIVRD--ITTNNELGRDVT 177  
Db 154 IDNNDLQGVGRIDCGAAASTLV---FNLANPT--TOKAPLILGDNVAVNGANGTLNVT 217  
QY 178 NNVAAPDYK----NVHNVYNADNSGPDVNVNIDFSKMKDYRDSIEVSRYSNGKSV-- 230  
Db 218 NGFIKVSRSKATVNVINI--GDQGIIMFTD-----ADNVNLN--LOANGATITF 265  
QY 231 ---DKWSQIPITFDKNKYAYLDTFEYK-----NGELHA--TGWNA----- 264  
Db 266 NGTDTGRLVLLSKN--AAATDENVTGSLGNLKIIEFTNVAVNGQLKANAGANAAGVIG 323  
QY 265 TNSALNYNEHEVILFDQTNGKEVAROEVRQSRPDVAKVYPO--VYGAANSNGENVFNI 322  
Db 324 TNGAGRAGEVVSVD--NGK--VA---TIDGQ-----VYAKDVLOSANA-----V 363  
QY 323 SDLDYTHOYQVLSRYSNDNGEGONVTYWFNPQSIAPANQOGLYDSDISKN-----GE 378  
Db 364 GOVNFRIHVDV-----GTDGTAFTAKTAAKVAITQNSFGTTDFGNLAAOIIVPT 414  
QY 379 VTVTGNWATDISELONNHYVILFDQTAGKQVAKADL-----ISRPVAKAYPTVKT 431  
Db 415 MTLNGFTGDASNPONTAGVTFD--ANGTILASADANAVTNNITAIEASGAGVQLS 472  
QY 432 ATNS-----GFKYTEKVN-----LQPGHOYSVYSRESADENG--GN 467  
Db 473 GTHAELRLGNAGSVFKLADGTIVNGKVNQALVGGALAG---TITLGGSATITGIGN 529  
QY 468 DKRHTDYWFSPVILNQATSNIDTITMTSNGLHAGWASDINSINETTPYAILNNGKEVT 527  
Db 530 AGG-----AAALOGITLANDATKTLTGGANIIG--ANGGTIN-----FOANGG--- 571  
QY 528 RQKMSLTARPDVAAYPSLYNSAVSGEDTTI-----KLTDQYQALNGLOVLL 576  
Db 572 --TIKLT-----STQNNIVDFDLATATDOTGVWDASSLNTAQTITINGKI----- 615  
QY 577 RFSKAADNPNGDNTVDQFSKNYATGGNFDYKVN---GNO--VEFSQWHAAT--NQ 627  
Db 616 -----GTVGANKTLGQFNIGSKTVLSDGDVAINELVIGNNGVQFA--HNTYLITR 666  
QY 628 SNDXDSOMIIVLVNGKEVKROLVNDTRKGAAGFENDYKVNPALE--NSSMSGFOGIIITL 686  
Db 667 TTNAGQ-----GKIIFNPVNNNTLATGTLGSA--TNPLAEINFGSKGAANDVT 717  
QY 687 PVTYKENVQLVHRFSNDVKTEGNYVDFWSELMPVKDSFOKNGPLKQFLOLQINGQY 746  
Db 718 LNVKGVNL-----YATNITTTDANVGSF-----IFNAG-----GTNIVSG--- 753  
QY 747 YIDPTGQPRKNFLQSGNNHIYFSDTGVCTNALELOFAKGTVSSNEQYRNGNAAYSYD 806  
Db 754 ---TVGG-----QQNKENTVALDNGTIVKFL-----GNATF--NGNTTTAAN 791  
QY 807 DKSIEENVGYLTADTWYRKPQILKDGTTWTDSKETDMRPLMWMPNTLT---QAYYLNY 863  
Db 792 --STLIQGNYTADF-----VASADGTGIVEFVNTG-----PITVLNKOAAPVNA 835  
QY 864 MKQ-----HGULLSALPFFNADADPAELNHYSEIVQONIEKRISSETGNTDMLTLHDF 918  
Db 836 LKQITVSGPQVNV-----INEIGNAG----- 856  
QY 919 VTNNPMWKNDSNVNFGIQFGGFLKYENSDLTPYANSDYRLILGRMPINIKDOTYRQE 978  
Db 857 -----NYHGAV-----TDTAFENSSLGAVVFLP-----RGIP 884  
QY 979 FLANDINDNSPVVQAEOLNMLYLLNFGTITANNDQANFD--SVRYDAPDNIDAD--LM 1034  
Db 885 F---NDAGNRIP-----LTIKSTVGNKTAGTFDVPVSVILVGVDSVIADGOVI 928  
QY 1035 NTAQDYFNAAGMDSDAYSNKHINLLEDWNHADPEY--FNKIGNPO--LIMDDTIKSLN 1090  
Db 929 GDQNNIVGLGSDNDIIVN-----ATTIYAGITPINNOGTVTLSGGIPIPTNG 977

QY 1091 --HGLSDATNRWGLDAIVHOSLA--DRENNSTENVIPNYSFVRAHDNNSQDQIONAIRDV 1147  
Db 978 TVYGLGT-----GIGASKFKQVTFFTDYNLGNIIATNAT-----INDGTVT 1020  
QY 1148 TGXDYHTFFEDQKGDIDAYIQDONSTVKYKYNLYNPASVAILTNKDTIPVYICD--- 1204  
Db 1021 TG-----GIAGIGFDGKITLGSVN-----GNGNV 1044  
QY 1205 LYTDGGQYMEHOTRYDYDTLNLKSRVKYVAGSQSMQTMVSQVGNNNILTSVYRKGAMTA 1264  
Db 1045 RFVDG--ILSHSTSMGT--TKANNGTVTYLG---NAFVGNIGSDPTVASVAF-----TG 1093  
QY 1265 TDTGTDETRTQIGVVVVSNTPNLKLGVNDKVLHMAAHKNOQ--YRAAVLITTDGVINY 1322  
Db 1094 SDGAGLQGNISQVDFGTYNLGIS--NSNVILGGGTTAINGKINURTLTFLFASGTSTW 1152  
QY 1323 TSDQGPVAMTDENGDLYLSSHNLV-----VNGKEEADA--VOGYANDPSVGYLAVV 1374  
Db 1153 GNNTSIETTLTLANGNI---GNIVILEGAQVNAITTTGTTIKVQDNANANFSGTQTYTL 1208  
QY 1375 PVGASNDQARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSNEVYPTTKESERANVRIAQ 1434  
Db 1209 IQGARGFNGTLGGPFNVVTG-----SNRFVNYGLIRAAQNDYVITRTNNAENVVINDIA 1262  
QY 1435 NADFFASLG---FTSEMAPQ-----YNSKDRTELDSTIDNGYAFTRDYDLGM 1480  
Db 1263 NSFFGAGPVGQNVVTFVNAATNAAYNNLLAKNSANSANFYCAIYTD7SAAITNAQLDV 1322  
QY 1481 SEPKNYGTDEDLRNAIOALHKAQLQWADWPQIYVNLPGKEVATVTRVDDRG--NWKDA 1539  
Db 1323 AK---DIOAQLGNRLGALRYLGTPTETAEMAGPEAGAIIPAAVAAAGDEAVDNVAYGIWAKP 1378  
QY 1540 IINNLYVYNTIGGGEYOKYGGAEFLDKLQYPELFTTKQVSTGVAID---PSOKITE 1595  
Db 1379 FYTD-----AHQSKKG---LAGYKAKTTGVVIGLITLANDNLMI 1415  
QY 1596 WSAKYFNGTNILHRSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQGNFVKGNDGN 1655  
Db 1416 GAAIGITKTDIKH-----QDYKKGDKT-----DYN 1440  
QY 1656 YFVYDLAG--NMVKNTFIEDSV-----GNWYFFEDQD---KMWENKHFVDVDSYG 1700  
Db 1441 GFSFSLYGAQLVKNFPAQSAIFSLNQVKNKSQRIFFDANGNMSKQIAAGHT--DNMTFG 1499  
QY 1701 EKTYFFLKNVSGFRGLVOTDNGTYFF-----DNY---GKMVRNQTINA 1742  
Db 1500 GLNTVGYDYN--AMQGVLTMPAGLSYLVKSSDENYKRTGTTVANKQVNS 1546

## RESULT 14

BIGASALTY  
ID BIGASALTY STANDARD; PRT; 1953 AA.  
AC P25927; P25928; Q9XCQ3;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative surface-exposed virulence protein bigA precursor.  
GN BIGA OR STM3478.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 14028;  
RA Stojiljkovic I., Valentine P., Heffron F.;  
RT "Salmonella typhimurium rhs homolog.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SCSCL412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;



```
QY 1085 IKNSLNHGLSDATNRWGLDAIV---HQSLADRENNSTENVVIPN---YSFVRAHD----- 1133
Db 1194 V--TINGANALANYGTLDADAISTWISLFNEADSGSITDILLTNGDVTYNNGDFTGSI 1251
QY 1134 --NNSQOIQNAIRDVT---GKDY--HTTFDEQKIDAYIQDQNSTVK--KYNYLNI 1183
Db 1252 AGTSYQOEIVNT--GDMTVAEDGKSLVSGSFYYNEE---DATLNSGSAVEGSENTIINL 1307
QY 1184 P-ASYAILLTNKDITPRVY-YGDLYTDGGQYMEHQTRYDITLN-----LKSRY 1231
Db 1308 TRANDSLTOVNSGITATNGYSAITTVNGS---NDPKIWIATATGVINGINPDAPLINLGR 1365
QY 1232 KYVAGGSMQTMVSGGNNILTSVRYKGGAWATDTGT-----DETQIGIVVVSIN 1283
Db 1366 GYNEGNOG--TINVQGDNAVAISGTSYVINLVNSGTINVTGQEKEDGTNGTLI--- 1420
QY 1284 TPNLKLGVNDKVVLLHMGAAHKNQOYRAAVLTITDGVINYSQO---GAPVMTDENGDL 1339
Db 1421 -----GI-----KGNATINTTADGVINVAADSYAPGGTKAIINNGEI 1462
QY 1340 YLSHNLVYNGKEBADTAVOGYANDPVSGYLVAVPVGASDNODARTAPSTEKNSGNSAY 1399
Db 1463 NL-----LCDSGCDIYAPGTTGTQN--DHNGTADIVIP-----DATTAP-TEGSIPTPPA 1509
QY 1400 RTNAAFDSNVIFEAFSNVYPTPTKESERA-----NVRIAONADEFFASLGTSEMAPQYN 1454
Db 1510 DPNAP-----QQLSNIIVGTNADGSSGTLKANNLVIGDNYK--VDTGFTS----- 1552
QY 1455 SKDRTELDSTIDNGYAFTRDYDLGMSEPNKYGTDEDLRNAIQALHKAGLVQVMAWVDPQ 1514
Db 1553 GTADTTVV---VDN--AFTG-----SNIQAGDNITSVS----- 1581
QY 1515 IYNLPGRVAVTVTRVDRGNVWKDAIINNLY--VVNTIGGGEYQKYGGAFLDKLOKLY 1572
Db 1582 VNAQSGO-----DADGNV--DVMTKRAYADVATDSSVSDVAQALDAGYINN----- 1627
QY 1573 PEITTKOVSGVADPSQKITE--WSAKYFNGTNILHRSGYVLYKADG-----GQYIN 1624
Db 1628 -ELYTSLNVGPTABELNSALKOVSGAQTTFREARVL--SNRETMLADAAPQIKDGLAFN 1684
QY 1625 LGTTTKQFLPIQLGKKGKQNEGVKNGDYFYFDLAGNM---VKNTEFDSVGVNWF 1680
Db 1685 V-----VAKGDPRELNDGTQYDMLARQLDLTASONTLEYGIAR--- 1726
QY 1681 FDQKQMVENKHFVDPSYGEKGYFFLKNGVSFRGLVQTDNGTYFDRN-----YG 1732
Db 1727 LDGSG-----SKTAGDNGLTGGYSQFFGLKHSMAFDEGLANNLSLRVDVHNLDSRSRYAG 1782
QY 1733 KM-----VFNQI---NAGAMITYIDENCKLIKA----- 1758
Db 1783 DVNKIADSMRQQXLEFRSEGAKTFTM--MGDAUKVTPYAGVKFRHPTMEDGYKERSAGDF 1840
QY 1759 --SYNSDAEYPTSTDVGKMLD 1777
Db 1841 NLSMNSGNETAVDSIVGLKLD 1861

RESULT 15
OMPBL_RICTY
AC P96989; STANDARD; PRT; 1645 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiellae; Rickettsia.
OX NCBI_taxid=785;
```

```
[1]
SEQUENCE FROM N.A.
RP STRAIN-Wilmington;
RC MEDLINE=94040787; PubMed=8224886;
RX Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
[2]
PARTIAL SEQUENCE.
RP STRAIN-Wilmington;
RC MEDLINE=92114896; PubMed=1370573;
RX Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
[3]
IDENTIFICATION OF CLEAVAGE SITE.
RP MEDLINE=92104668; PubMed=1729180;
RX Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC !- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC !- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC !- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC !- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L04661; AAB48987.1;
DR InterPro; IPR003858; rOmpA_rOmpB;
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1354 1645 32 KDa BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 637 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 2.8%; Score 264.5; DB 1; Length 1645;
Best Local Similarity 19.8%; Pred. No. 8.8e-05;
Matches 396; Conservative 230; Mismatches 729; Indels 643; Gaps 94;
```

Qy	234	SQITFDKN-----NYAVL---DTEFVKNGELHATGNWATNSAINYHNHHVFILFDQUNG	268
Db	179	AAPSVLEFNLIINTPTDEAPLITGDNAKIYNG-----ANGILNITNGFVKVSDKT--	227
Qy	285	KEVARQVREGRQSRPVAKYIPVQVGAAN---SGENVTFNISDLDYTHQYQVLSRYNSD	341
Db	228	-FAGIKTINIGDNOGLMFWTTPDAANALNLOGGNTINFNGRD-----	269
Qy	342	NGEGDNVYWFNQSTAPANQSNQGYLDSFDISKNGEVT---VTGNWATDISELQNNHYV	398
Db	270	-GTGKLVL-----VSKNGNATEFNVTSQGLGNLKG-----V	299
Qy	399	ILEDOT--AGKQVASAKADLIISRPDVAKAYPTVKTATNSGFKVTFKVNLPQGHQYSVVS	456
Db	300	IEFDTTAAAGKLIANGA-----ANAVIGTDNGAGRAAGFIV-----	336
Qy	457	RFSADENGNGDKRHDTYFSPVILNOTASNTDITMTSINGLHIAGWASDNSINETPY	516
Db	337	--SVD-----NGNAATISQGVAKDIVTOSA-----	360
Qy	517	AIIILNNGKEYTRQKMSLTARPDAAYVPSLYNSAVSGFDTIKLTNDQYQALNGLOVLV	576
Db	361	---NAGQVTFEHL-----VDVGLGKTKFKTADSKVLIITENASFGSDTGNLAVQIV	411
Qy	577	RFSKAADNPSGD-----NTV--TDQFSKNYATIGNFD--YVKVNGNOVFSGHWATNQ	628
Db	412	PNNKILFTNGIGDAKNNGTAGVITFENANGTILVSGTDPNIVVTWIKAEIEGAGIVQLS	471
Qy	629	NDKDSQWIIVLVNGKEYVKQLVNDTKEGAAGFNNDVKVNPATENSMSFGQIITLPV	688
Db	472	G-----IHGAELRLGNAGSIFKLADGVINGVQNQLPVNNAALAA--GSIQLD--	518
Qy	589	TVKNENVQLVHRSNDVKYTGEGNVYDFWSELMPVKDSQKGNPLQKPGLOTINGQOYVI	748
Db	519	-----GSAIITGD-----INGAVNA--ALQ-----	537
Qy	749	DPTTGOQPKNFLLQSGNNMIYPDSDTGVGTN--ALELQFAKGTVSSNEQYRNGNAAYSV	805
Db	538	DIITLANDASILLTSGANII-----CANAGGAIHFQANGGIQILTS--TONNILVDF	587
Qy	806	DDKSIEENVGYLADTWYRPKQLKDGTTWTDQSKETDNRPILMVWPNLTFOAYLYNMYK	865
Db	588	DLDTTDDTQGVDPASSLTNNQTTITSGITIGANTKILGRFNVGSSKTIUNA-----	640
Qy	866	QHGNLLPSALPFENADAPAEALNHYSIEVQOONIEKR-----TSETGNTDQLTLMHDF	918
Db	641	--GDVAINELVMENDOS--VHJTHNTYLTITKTINAAANGKIIVAADPTIND--TALAD--	692
Qy	919	VTNPNMNKDSENVNFSGIOGGFLKYENSOLTPYANS-----DYRLLGPMPI	967
Db	693	GTNLGSAESPLSIHFATKAAANGDSILHIGKGVNLYANNITTTDAWVGSGLHFRSGG--T	749
Qy	968	NIKDQYRGQEFFLLAND--IDNSNPVVOAEOLNWLILNFGTITANN-----	1014
Db	750	STVSGTVGGQGGUKLNLILDNGTV-----KFLGDTFNGGTKIEGKSILQI	797
Qy	1015	QANFDSVRVDAPDN-----IDADLMNIAQDYFNAAYGMSDVAENKHNILEDWNHADP	1068
Db	798	SSNYITDHIIESADNTGLEFVNTDPTITVLNKGAVFGVLQKQVMVSGPNTA-----	849
Qy	1069	EYFNKTGN---POLTMDD-----TIKNSLNHGLSDATNR---	1099
Db	850	--FNEGTNGVYAHAIYDVISFENASLGASFLFLSGTPLDLVLTIKSTVGVNGTVDNFAPIL	907
Qy	1100	--WGLDAIVHQSLADRENNSTENVIPNYSFVRHDDNSQOIQNAIRDVTKGYHYHFTF	1157
Db	908	VVSSIDSMI-----NNQ-----VIG-----	923
Qy	1158	EDQOKGIDAVI--QDONSTVKKYNYNTPASVAILLTKNDTI-----PRVYGDLY	1206
Db	924	--DQKNIIALSLGSDNSITVNSNIIY--AGIRTKTKQGTWTLGGIPNPNPFIYQ--LG	977

Search completed: April 23, 2003, 15:08:21  
Job time : 61 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 15:03:26 ; Search time 70 Seconds  
(without alignments)  
5242.427 Million cell updates/sec

Title: US-09-995-749A-2

Perfect score: 9424

Sequence: 1 MEKKHFKLYKSGKWTA.....SDAEYPTSDVGKMLDQNK 1781

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2724	28.9	2057	Q9RE05	Q9RE05 leuconostoc
2	2546.5	27.0	1527	Q9ZAR4	Q9ZAR4 leuconostoc
3	2490.5	26.4	1477	Q9L466	Q9L466 leuconostoc
4	2483	26.3	1508	Q52224	Q52224 leuconostoc
5	2479	26.3	1508	Q9EZH5	Q9EZH5 leuconostoc
6	2375.5	25.2	1577	Q54178	Q54178 streptococc
7	2373.5	25.2	1449	Q68542	Q68542 streptococc
8	2373.5	25.2	1449	Q55264	Q55264 streptococc
9	2346.5	24.9	1455	Q69397	Q69397 streptococc
10	2343.5	24.9	1455	Q69388	Q69388 streptococc
11	2342.5	24.9	1455	Q69382	Q69382 streptococc
12	2341.5	24.8	1390	Q69385	Q69385 streptococc
13	2340.5	24.8	1455	Q69391	Q69391 streptococc
14	2330	24.7	1575	Q9LCH3	Q9LCH3 streptococc
15	2289	24.3	1590	Q55263	Q55263 streptococc
16	2284	24.2	1590	Q59983	Q59983 streptococc

17	2256	23.9	1290	2	Q48756	Q48756 leuconostoc
18	2190.5	23.2	1577	2	Q55265	Q55265 streptococc
19	2190	23.1	1338	2	Q9WXJ4	Q9WXJ4 streptococc
20	2119.5	22.5	1512	2	Q9WXJ5	Q9WXJ5 streptococc
21	2093	22.2	1518	2	Q00600	Q00600 streptococc
22	2005.5	21.3	1599	2	Q00599	Q00599 streptococc
23	1989	21.1	1016	2	Q9LCJ7	Q9LCJ7 leuconostoc
24	764.5	8.1	591	2	Q8VUH3	Q8VUH3 streptococc
25	748	7.9	522	2	Q8VV10	Q8VV10 streptococc
26	424	4.5	2817	16	Q97K42	Q97K42 ciosiridium
27	326.5	3.5	2462	16	Q8RGZ3	Q8RGZ3 fusobacteri
28	326.5	3.5	2806	16	Q8RI19	Q8RI19 fusobacteri
29	321.5	3.4	2353	2	P71401	P71401 haemophilus
30	321.5	3.4	3165	16	Q8RDQ9	Q8RDQ9 fusobacteri
31	321.5	3.4	4688	16	Q9PQ08	Q9PQ08 ureaplasma
32	318.5	3.4	6713	16	Q99U54	Q99U54 staphylococ
33	317.5	3.4	2276	2	Q93TY6	Q93TY6 staphylococ
34	317.5	3.4	6713	16	Q93IR6	Q93IR6 staphylococ
35	317	3.4	2399	16	Q9ZKS9	Q9ZKS9 helicobacte
36	315.5	3.3	2529	16	Q25579	Q25579 helicobacte
37	313	3.3	2902	16	Q9ZMB6	Q9ZMB6 helicobacte
38	311.5	3.3	1612	16	Q8Y591	Q8Y591 listeria mo
39	311.5	3.3	4919	2	Q9ZHL0	Q9ZHL0 haemophilus
40	309.5	3.3	2217	17	Q8TP72	Q8TP72 methanosarc
41	309	3.3	2893	16	Q25063	Q25063 helicobacte
42	308	3.3	3194	16	Q9ZLM3	Q9ZLM3 helicobacte
43	308	3.3	3269	2	Q9RGN5	Q9RGN5 lactobacill
44	305.5	3.2	1806	16	Q929J3	Q929J3 listeria in
45	304.5	3.2	1463	2	O86919	O86919 staphylococ

ALIGNMENTS

RESULT 1

Q9RE05 PRELIMINARY; PRT; 2057 AA.  
ID Q9RE05  
AC Q9RE05  
DT 01-MAY-2000 (TREMELrel. 13, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)  
DE Alternansucrase (EC 2.4.1.140).  
GN ASR.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Leuconostoc.  
OX NCBI\_Taxid=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1355;  
RX MEDLINE=20080809; PubMed=10612736;  
RA Arguello-Morales M.A., Rемаud-Simeon M., Pizzut S., Sarcabal P.,  
RA Willemot R.M., Monsan, P.;  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
RL FEMS Microbiol. Lett. 182:81-85(2000).  
DR EMBL; AJ250173; CAB65910.2; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; GH\_70.  
DR Pfam; PF01473; CW\_binding\_1; 12.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Glycosyltransferase; Transferase.  
SQ SEQUENCE 2057 AA; 228987 MW; 62BCE9385D9A11BE CRC64;  
Query Match 28.9%; Score 2724; DB 2; Length 2057;  
Best Local Similarity 36.8%; Pred. No. 1.7e-100;  
Matches 677; Conservative 210; Mismatches 453; Indels 500; Gaps 56;

QY 8 KLYKSGKWTAAV--ATVAVSTALLYGGVAHADQQVQAASQTQDSTVNNDTKTVAL 65  
Db 11 KLYKSGKWVAATAFAVLGVSTV---TVHADTNSNVAVKGINNTGTNDSEKQVVP 66  
QY 66 DTNTDQSAQTDRKQVVSNTNQ--SKTDD--TSTADKNSTSTPVSVLPSNTEKQAKNY 120



```
Db 67 STNDSLKQGTGDFWYSDGSRVQDKTNQILLTAELKKNNEKNLSVISDDTSKDDENI 126
QY 121 NEQDK-GNYGNIDTAYFSNNQLHVSNGWATNASQTSRQIIIVRIIDITNNELGRTDVTNN 179
Db 127 SKQTKIANQCVDTA-----KGLTTSN-----LSDP 152
QY 180 VARDVKVNHVY--NADNSGFDVN--VNIIDFSKKKYRDSIEIVSYSGNGSKVDWMSQ 235
Db 153 ITGGHYEN-HNGYEVYIDASKQVTLQNIID-----GNLQ----- 186
QY 236 PITDKNKYAYLDTFEVKNGELHATGNWATNSAINYNHHFVILFDQNGKEVARQEVREG 295
Db 187 --YEDDNGYQVGSFRVNGK-----H--IYFDSVTGRASSNDVIANG 225
QY 296 QSRPDVAKVYPQVGAANSRGNVTINISDLDTYQYVLSRYNSDNGEGNVTWYFNPQ 355
Db 226 KAQ-----GYDAQGN-----OLKKSIVADSSGQ-----TYFED-- 253
QY 356 SIAPANQSNQYLDSDISKNGEVTVTGWNATDLSELQNNHYVILFDQTAGQVASAKAD 415
Db 254 -----GNQ----- 257
QY 416 LISRPDVAKAYPTVKTATNSGFKVTFKVNLPQGHQYVSVRFSADENGNGNKRHTDYW 475
Db 258 -----PLIGLOTIDG-----NLQYFNOQGVQIK-----GGFQDVNNKRIY 292
QY 476 FSPVILNQTASNIDITWTNGLHAGWASDINSINETTPVAIILNNGKEVTROKMSLTA 535
Db 293 FAP-----NTGNNAVANTEI----- 306
QY 536 RPDVAAYPSLYNSAVSGFTTIKLTNDQYQALNGQQLVLRFSKAADGNPSGNTVTDQ 595
Db 307 -----INGKLQ-----GRDANGNQVNA 324
QY 596 FSKYATTTGGNFDVYKVNQVFEFSGWHATNQSDKDSQWIIIVLNGKEVKRQLVNTKE 655
Db 325 FSKDVA-----GNTFVFDANG----- 340
QY 656 GAAGFNRDVYKVPAINESMSGFGIITLPVTVKENVOLVHRFSNDVKTGSGNVDF 715
Db 341 -----VMLT----- 344
QY 716 WSELMPVKDSQKNGPLKQFGLQTINGQOYYIDPTTGPRKNLQSGNNWYIFDSDTG 775
Db 345 -----GLQTSIKTYILD-EQHLRKNVAGTFNNOFYFADATG 382
QY 776 VGTNALEQFAKGTYSSEQYRNGNAAYSDDKSTENYNGYLTADTWYRPRQILKDGTTW 835
Db 383 AGKTAIEYQFQDGLVQSQSNENTPNAAKSYDKSSFENVQGYLTADTWYRPTDILKNGDTW 442
QY 836 TDSKETDRPILWVWPTLTQAYLYNKKHGLLPSALFFFNADADPAELNHYSETVQ 895
Db 443 TASTETDRPILMTWPDQTKQANYLNFMSKGLGITTT---YTAATSQKTLNDAARFVQ 499
QY 896 QNIEKRISGTGTNIDLRLTMHDFVTNNPWKNKSDSENVAFSGIQF-QGGFLAY-ENSLDTP 953
Db 500 TAIEQOISLKKSTELWRDAIDSVFKYQANWKNQTEDEAFDLQWLOGGFLAYQDDSHETP 559
QY 954 YANS-DYRLLGRMPINI---KDQT-YRQGEFLANDINDSNPPVQAEOLNWLXYLLNPGT 1008
Db 560 NFDSGNNRKLGRQPINIDGSKDITDGGSEFLANDINDSNPIVQAEOLNWLHYLMNFGS 619
QY 1009 ITANNDQANFDSVRVDPADNIDADILMNTAODYENAAQGMQ-SDAVSNKHINILEDQNHAD 1067
Db 620 ITGNNDNANFGIRVDVNDVADLILKTAGDYFRALYCTKSDANANKHLSILEDWNGKD 679
QY 1068 PEYFNKIGNPQITMDTIIK-----NSLNHGLSDATNRW-----GLDAIVHQSLADREN- 1115
Db 680 PQVYVNOQNAQLTMDYTYVTSQFNSLTHGANNRSMWYFLDTGYVYNGDLNKKIVDKNR 739
QY 1116 -----NSTENVIPNVSFVPAHDNNSQDQIQNAIRD--VTGKDYHFTFEDEQKG 1163
Db 1116 -----NSTENVIPNVSFVPAHDNNSQDQIQNAIRD--VTGKDYHFTFEDEQKG 1163
```

```
Db 740 NSGTLVNRANSODTKVIPNYSEVRAHDYDAQDPIRKAMIDHGIILKNMODTFTFDQLAQG 799
QY 1164 IDAYIQDN--STVKKYLNLYNPASVAILLTNKTIDPRVYVYGLYTDGGQYMEHQTRYD 1221
Db 800 MEFYFKQDNPSGFKKYNDYNLPSAYAMLLTNKDTVPVYVYGMVLEGGQYMEKGTIYNP 859
QY 1222 TLTNLLKSRVYVAGQSQMOTMSVG-----GNNNLTISVRYGKGAMTATDCTDET--- 1272
Db 860 VISALLKARIKYVSGQTMATDSSGKDLGDETDLTISVRFKGIMTSQDTITQNSQDY 919
QY 1273 RTGIGGVVSYNTPNLKLGVNDKVLHMGAAHKNQYRAAVALTTDGVINYSDDQAGPAM 1332
Db 920 KNOGICVIIVGNPDLKLNNDKIITLHMGRAHKNQLYRALVLSNDGIDYVDSDDKAPTLR 979
QY 1333 TDENGDIYLSHNLVNGKEEA---DTAVQGYANPDVSGYLAWVYPVGSADNQDARTAPS 1389
Db 980 TDNGDLIIFHKNTFY--KODGTIINEMKGSUNALISYLGWVYVPGASDSODARTV-A 1036
QY 1390 TEKNSGN--SAYRTNAAFDSNVIFEAFSNFVYPTTKESERANVRIAQNADFFASLGFTSF 1447
Db 1037 TESSSSNDGVSFHSNAALDSNVIEGFSNFQAMPTSPEQSTNVVIATKANLFKELGITSF 1096
QY 1448 EMAPQYNSKD-----RTFLDSTIDNGYAFTDRYDLGMSE-----PNKYGTDEDLRNAI 1496
Db 1097 ELAPQYRSSGDTNYGMSFELDSFLNNGYAFTDRYDLGNKADGNPNPTKYGTDDQLRNAI 1156
QY 1497 QALHKAGLQVMAQVDPQIYNLPQKEVATVTRVDDRGNTWKDAIINNLYVYNTIGGG-E 1555
Db 1157 EALHKMGQAIADWPDQIYALPGKEVATVTRVDERGNOLKDTDFVNLVYVANKSSGVD 1216
QY 1556 YQKKYGAFLDKLQKLYPEIFTKKQYSTGVAIDPSOKITEWSAKYPNGNINLHRSGYVL 1615
Db 1217 YQAKYGEFLDKUREEYPSLFQONQVSTQPIDASTKIKQWSAKYMNNGTNILHRSYVYL 1276
QY 1616 KA-DGGYVYNLGTTTKQFLPIQLTGEKKQNGEFGVKGNDGNYFYDLAGNMVKNKTIEDS 1674
Db 1277 KDWATQYFNIAKTNEVFLPLQL--QNKDAQIGFISDASVKY-YISGYSQAKDIFEDG 1333
QY 1675 VGNWYFDDQGGKVVENK-----HFVDVDSYGEKGYFFLKNGVSPRGLVQTDN--GYI 1726
Db 1334 NGWYIFDKDGYVVRSGQGNPIRTVETSVNTRNGNYFYMPNGVELRKGF-GTDNSGNYI 1392
QY 1727 YFNYGKVMKVNQOTINAGA-MIYTLDENGKLKASYNSDAE 1765
Db 1393 YFDDQGNWRDCKYINDDANNFYHLNVDGTMRSRGLFKFEDSD 1432
RESULT 2
Q9ZAR4 PRELIMINARY; PRT; 1527 AA.
ID Q9ZAR4 AC Q9ZAR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Dextranucrase.
GN DEX.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Leuconostoc.
OX NCBI_Taxid=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512-F;
RA Bhatnagar R., Singh D.K.S.;
RT "Cloning and Molecular Characterization of Dextranucrase Gene from
RL Leuconostoc mesenteroides NRRL B-512F." ;
DR Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U91374; AAD10952.1;
DR InterPro; IPR002479; CW_Binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_Binding_1; 16
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;
```

```
Query Match      27.0%; Score 2546.5; DB 2; Length 1527;
Best Local Similarity 43.8%; Pred. No. 1.4e-93;
Matches 578; Conservative 178; Mismatches 405; Indels 159; Gaps 34;

QY 537 PDVAA--YPSLYNSAVSGFTTIKLTNDQYQ--ALNQLQVLLRFSKAADNPSGDNV 592
DB 47 PDVSANNYQASDNTDTTQQNTVTTEENDKVKQSAATNDNVT-----TAASDTTQASDNNV 101
QY 593 TDQFSKNATATGGNPDYKVGNGNOVEFGSGWATNQS-----629
DB 102 TEKQSDHA---LDNEKVDNKQDEVAQTNVTSKNEESAVASTDTPAETITDETQVCS 156
QY 630 ---DKDSQWIIVLVNGKEVRQLVNDTKEGAAGFNNDVYKVPALNENSMGSGFQGIT 685
DB 157 GXYEKGDSWYFFPDGNKAK-----GLSTIDNNIQY----FYESGQKAGQ-----199
QY 686 LPVTYKNNVQLVHRFSNDVKTG-----EGNVYDFWSELMVP-KDSFQKNGPLKQF-----736
DB 200 -YVTDNQTYYF-DKSGDELGTGLQSIDGNVAFNDEQQIFNQYQSENGTYYFFDCKG 257
QY 737 ---GLOINGQYIIDTTGQPKNFLLQSGNNWIFDSDTGVGTNALEQFAKGTVSS 792
DB 258 HAATGKIKIEGKNYFD-NLGOLKKGFGSGVIGDQIMTFDQETGQEVSNVTTSEIKEGLTQ 316
QY 793 NEQYRNGAAYSDYDKSIENYNGYLTDATWYRPKQILKDGTTWTDSKETDMRPLMWWP 852
DB 317 NTDYSEHNAAHGCTDAEDFENIDGYLTASSWYRPTGILRNGTDWEPSTDTDRPILSWP 376
QY 853 NPLTQAYLYNKQHGNNLLPSLPFNADAPAE-----LNHYSEIVQONIEKRISSET 905
DB 377 DKNTQVNYLYNADLG-----FISNADSFETGDSQSLNEASNVYQKSIEMKISAQ 427
QY 906 GNTDRLRLTMHDFVNTNPMWKNKDSNVNFGSIQFOGGLKYENSDLTPYANSVRLGLRM 965
DB 428 QSTEWLKDAMAFIVAQPOWNETSDM--SNDHLQNGALTYVNSPLTPDANSFELLNPT 485
QY 966 PYN-KDQTYR-----GOEFLANDIDNSNPVQAEOLNWLNYLLNFTGTIANQDQANF 1018
DB 486 PTNQTGEQAYNLNDSKGGFELLANOEDNSNVVVEAQLNMLYLLMNFGTITANDADANF 545
QY 1019 DSVRYDAPDNIDADLMNTAQDFYNAAYGMD-SDAVSKHNILIEDNWDHAPFEYKNGP 1077
DB 546 DGIYDAVDNVDADLLQIAADYFKLAYGVQNDATANGHLSILEDHSHNDPLIYTDQGSN 605
QY 1078 QLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSADRENNTENVIPNYSFVRAHDNNS 1136
DB 606 QLTMDDYVHTQLIWSLTKSSDIRGTMRQFVDYVYMDRSDNSTENEAIPNYSFVRAHSEV 665
QY 1137 QDQIQNAIRDVTGKYHFF--TFDEBQXGIDAYIQDQNSTVKKYNLYNIPASYAIIILTNK 1194
DB 666 QTVIAQIVSDLYPDYENSILAPTEQLAAAFKYVNEDEKLADKKYQYNWASAYAMLLTNK 725
QY 1195 DTPRVYVGDLYTDGQVMEHTRYDYTLNLLKSRVYAGGQSQMOTMSGVGGNNILTS 1254
DB 726 DTPRVYVGDLYTDGQVMAKSPYDYAINTLLKARQVYAGGQSNVSDS-----NDVLTS 781
QY 1255 VRYGKAMTATDGTDETRTQGIQVGVNSTPNLKLGVNDKVVLLHMGAAHKNOQYRAAVIT 1314
DB 782 VRYGKAMTASTGTSEITRTEIGIVSNNAELQLEDGHTVTLHMGAAHKNOQYRALLIST 841
QY 1315 TTDGVINYTSDQCAPVAMTDENGDIYLLSHNLVNGKREADTAVQGYANPDVSGYLAVMY 1374
DB 842 TADGLAYYTDENAPVAYTDANGDLIFTE-----SIYGVQNPQVSGYLAVMY 889
QY 1375 PGCASNDQARTAPSTEKNSGNSAVRTNAAPDSNVITFEAFSNFVYTPKESERANVTAQ 1434
DB 890 PVGAQDQDQARTASDTTNTTSKVFHSNAALDSQVIYEGFSNFQAFAPADSSEYTNVNTAQ 949
QY 1435 NADFFASLGTSPFMAPQNSKSDRTFLDSTIDNGYAFTRDYLGLNSENPNKTDDEDLRN 1494
DB 950 NADQFKQGWGVSFQLAPQYRSSTDSFLDSIIQNGYAFTRDYLGVGTPKVTGTADQLRD 1009
```

```
1495 AIALKHAGLQVWADVPDQIYNLPKEVATVTRVDDRGNVWVDALINNLLVYVNTIGG 1554
1010 AIALKHAGLQVWADVPDQIYNLPKEVATVTRVDDRGNVWVDALINNLLVYVNTIGG 1069
1555 EYCKYGAFLDKLQKLYPFIETKKQVSTGVAIDPSQKITENSAKYFNENGTNILLHRSGYV 1614
1070 QYQEMYGGAFLDLQKLYPFIETKKQVSTGVAIDPSQKITENSAKYFNENGTNILLHRSGYV 1129
1615 LKADG-QQYVNLGTTK--QFLPQLTGEKKQGNFVKGNDGNYFYFDLAGNNVNTFI 1671
1130 LKMGSNKYEKVVSNTEEDGYLPKQLTNDLSE--TGFTHDKGIIY-YTLGSGYRAQNAFI 1186
1672 EDSVGNMYFDDQGMVE-----NKHFVDVDSYGEKTYFFLKNGVSFSGGLVQTONGT- 1725
1187 QDDNNYVYFDKTHLVTLGLOKINN-----TYFFLNGIELYKSFQNEEDGTI 1235
1726 YFEDYKMYVRNQTI---NAGMIYTLDEKGLIK---ASYNDAEYPTSTGVKMLDQN 1779
1236 YFEDKKGHVFQDQVITDQNGNA--YFDDAGVMLKSLATIDHQY-----FDQN 1284

RESULT 3
Q9L466 PRELIMINARY; PRT; 1477 AA.
AC Q9L466;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN DSRG.
QS Leuconostoc mesenteroides.
QC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
QX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NERL B-1355;
RA Arguello-Morales M.A., Renaud-Simeon M., Pizzut S., Sarcabel P.,
RA Willenot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ250172; CAB76565.1;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 14
DR Pfam; PF02324; Glyco_Hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1477 AA; 164887 MW; 6GF5710DEDFCB831 CRC64;

Query Match      26.4%; Score 2490.5; DB 2; Length 1477;
Best Local Similarity 41.9%; Pred. No. 2.3e-91;
Matches 560; Conservative 187; Mismatches 439; Indels 149; Gaps 36;

QY 491 ITMTSNGLIAGWASDNSINETTPYAILNNGKEVTRQ-----KMSLTARPD-----V 539
DB 2 LSMATQSNV-----NADSTNTVTDKSVTNSNNTTQHDVTVDKQITPIVKNDQTQQI 56
QY 540 AAVPSLYNSAVSGFTTIKLTNDQYQALNGLOVLLRFSKAADNPSGDNVTDQFSKN 599
DB 57 AA--NAQAEVKASDIT--TDQKQA-----ETANNTKDSIDNLTQKLPV 100
QY 600 YATTCGNFYDYKVGNGNOVEFGSGWATNQS-----DKDSQWIIVLVNGKEVRQLV 650
DB 101 TPTANQKTGYLEKDGK-----YVYVTSNDTLAKGLTVDNHNKQYFD--NNGVQAKGQFV 152
QY 651 NDTREGAAGFNNDVYKVPALNENSMGSGFQGITLTPVYKNNVQLVHRFSNDVKTGEG 710
DB 153 TDNSK-----TYLDPNSGN-AVTGQQIGSQTLAF-NDN-----GEQ 188
QY 711 NVDFSWSELMVPKDSF---QKNGNPLKQFGLQITNGQQYIIDPTTGQPKNFLLQSGNNW 767
DB 189 VFADFYT--APDGKTYFFDDKQATI---GLKAINGHNYFD-SIGQLKKGFTGVIDQV 242
```

```
Qy 768 IYFSDTGVGTNALELOFAKGTSSNQYRNGNAAYSDYDKSIENVNGYLTADTWYRPKQ 827
Dy 243 IYFDESQOEVTSDSOIKELISQTTDYTAHNAHVHSTDSADDFDNFNGYLTASSWYRPKD 302
Qy 828 IYDGTWTWSDKEDMPILMWPNLTQAYLYNYKQKHGNLLPSALPFENADADAPAE 887
Dy 303 YLRNGQWHEATANDFRPIVSWWPSKOTQVNYLYNYSQMG--LIDNRQMFSLKDNQAML 360
Qy 888 NYHSEIYQONIEKRISSETGNRTDMLRTHLHDFVTNPNMKNKDSNVNPSGIFOGGFLKYE 947
Dy 361 NIACITVQQALETIGVANSTAWLKTATDIFRTQPOWNMSSEDP--KNHDLQNGALTFV 418
Qy 948 NSDUTPYANDSYRLGLRMPIN-----IKDQYRGGEFLLANDIDNSNPVQAEQLNWL 1000
Dy 419 NSPLTPDNTNSFRLLNPTPTNQTGVPXYITDQSKGGFELLANDVDNSNPVQAEQLNWL 478
Qy 1001 YLLNFGTITANNDQANFSDVRVDAPNDIDADLMNIAQDYFNAAAGMD-SDAVSNKHINI 1059
Dy 479 HYLNFSGITANDSAANFDGIRVDVNDVADLLOIAADYFKAAYGVKDKNDATANQHLST 538
Qy 1060 LEDWNHADPEYFNKIGNPOLTMDDTIKNSLNHGLS-DATNRWGLDAIVHOSLADRENNST 1118
Dy 539 LEDWNSHNDPEYKVDGNNQNTMDDYMHMTQLIWSLTKDMRMRGTMOREMDYVYLVNRNHDST 598
Qy 1119 ENVVLPNYSFVRAHDNSODIOINARIDV--TGKDYHTFTFEDEQKIDAYIQDNSTVK 1176
Dy 599 ENTAIPNYSFVRAHDSEVQTIVAOIISLHLPDVKNSLAPTADQALAEAFKIYNDEKQADK 658
Qy 1177 KYNLYNPASVAILTNKRTIPRYVYDGLYTDGQYMEHQTRYDYDTLTLNLLKSRKVKYVAG 1236
Dy 659 KYTOYNMPSAYAMLLTNKRTIPRYVYDGLYTDGQYMANSPYFDAINGLLKSRKIKYVAG 718
Qy 1237 GQSQWTHSVGGNNILTSVRGKAMATATDGTDETRTQIGVGVVNTPNLKLGVNDKVV 1296
Dy 719 GQSMAY-----DONDILTNRVYRGKAMSVTDSGNADTRTQIGVGVSVKNENLAKSGDTVT 774
Qy 1297 LHMGAHKNQCYRAAVLTTPDGVNYSODGAPVAMTDENGDLTSLSHNLVYNGKEEADT 1356
Dy 775 LHMGAHKNQAFRLLTGTADNLSYDND-NAPKYINDQGLDIFD-----NT 821
Qy 1357 AVQGVANPDVSGYLAVVPVGASNDQARTAPSTSEKNSGNSAYETNAAFDSNVIFEAFSN 1416
Dy 822 EIVGVRNPQVSGFLAVVPVGASDQARTLSDDTAHDGKTFHNSAALDSQVIYEGFSN 881
Qy 1417 FVYPTRESERANVRIOAQNDEFASLGTSFEMAPQYNSKSDRTFLDSTIDNGYAFTRDY 1476
Dy 882 FOAATNTEDYTNVIAKNGOLFWDGILTSFOLAPQYRSSTDSFLDSITONGYAFTRDY 941
Qy 1477 DLGSEPNKYGTDEDLRNAIOALHKAQLQVWADWVPQIYNLPKGEVATVTRVDDRGNW 1536
Dy 942 DLGYTPTKYGTVDQLDAIKALHANGIOALADWVPQIYNLPQOELATVTRTNSYGDKD 1001
Qy 1537 KDAIINNLYVYNTIGGGEYOKYGGAFDLKLOKLYPEITFKQVSTGVAIDPSOKITEW 1596
Dy 1002 TNSDIDQSLYVQSRGGKYQAQYGGAFSLDIOKKYPALPETKQISTGLPMDPSOKITEW 1061
Qy 1597 SAKYFNGNTHLRSGSVYVLKADG-QQYNNL--GTTTKOFLPIQLTGKKQGNBGFVKGND 1653
Dy 1062 SKKYFNGSNIQCKGAGYVLKDSGTQYKYVTNSNNRNDLFPKQLTDDLSE--TCGFVRDNI 1119
Qy 1654 GNYFYDLIAGNMVKNTRFEDSVGNWYFDDQGVKNENKHFVDVDSYGEKGTFFELKNGVS 1713
Dy 1120 GMVY-YTUSGLYARNTFIQDNGNYFYFDSGTGHLVTG--FQNIINN----HYFFLPNGIE 1172
Qy 1714 FRGGLVQ-TDNGTYVFNQYKWRNQTI-----NAGAMIYT-----L 1750
Dy 1173 LVQSFLOWADNGSTIFYDQKGRQVFNQYITDGTGYAYFQNDGTWVTSGETIDGHKQYFY 1232
Qy 1751 ENKGLIKASYNSDAE 1765
Dy 1233 KNGTQVKGQFVSDT 1247
```

## RESULT 4

```
OS2224 PRELIMINARY; PRT: 1508 AA.
ID O52224
AC O52224;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glucosyltransferase (EC 2.4.1.5).
GN DSRB.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL B-1299;
RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;
RT "Cloning and sequencing of a gene coding for an extracellular
RT dextranucrase (DSRB) from Leuconostoc mesenteroides NRRL B-1299
RT synthesizing only a (1-6) glucan.";
RL FEMS Microbiol. Lett. 0:0-0(1998).
DR EMBL; AF030129; AAB95453.1; -.
DR InterPro; IPR002479; CW_Binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;
```

Query Match 26.3%; Score 2483; DB 2; Length 1508;  
Best Local Similarity 41.7%; Pred. No. 4.6e-91;  
Matches 563; Conservative 188; Mismatches 438; Indels 162; Gaps 38;

```
Qy 477 SPVLNQTASNDITMTTNGLHAGWASDMSINETTPYAILNNGKRVTRQ-----K 530
Dy 28 STIMLSPD-----TSQNVNA-----DSTNTVTDKSVTSNNSNTTNOHDTVYDK 72
Qy 531 MSITARPD-----VAAVPSLYNSAVSGEDTTIKLTNDQYQALNGLOQLLRFSAADGN 585
Dy 73 QIPVANDOTTOIAA--NATQAEKVKASDIT--TDIQKA-----ETANNT 115
Qy 586 PGSD--NVTVDQFSKNYATGGNFDVKVNGNOVEFSGWHAATNQSN-----DKSQ 634
Dy 116 NKEDSDNLTKQLPAVTPPTANOKTGYLEKDGK-----WYVTSNDTLAKGLTVDNHQ 169
Qy 635 WIIVLVNGKEVKQLVNDTKEGAAGNRNDVYKVPAINESMSGFOGIIITPVTVKEN 694
Dy 170 YFD--NNGVQAKQFVTDNSK-----TYLDPNSCN-AVTGLQIGSOTLAF-NDN 216
Qy 695 VQLVHRFSNDVKTEGNYVDVFWSELMPVKDSF---QKNGPLKQFGLQTINGQQYIDPT 751
Dy 217 -----GEOVFADFT--APDGKTYVFDKGOATI---GLKAINGHNYFD-S 257
Qy 752 TGQPRKNELLQSGNNWIFYDSDTGVGTNALELOFAKGTVSSNEQYRNGNAAYSDDKSTE 811
Dy 258 LGQLKKGFTGVIDGVRYFDQSGGQVSTVDSQIREGLTSQITDYTAHNAVHSTDSADF 317
Qy 812 NVNGYLTADTWYRPKQIKLDGTTWTDKETDRPILMWPNLTQAYLYNYKQKHGNLL 871
Dy 318 NFNGYLTASSWYRPKQVLRNGQWHEATTANDFRPIVSWWPSKOTQVNYLYNYSQMG--L 375
Qy 872 PSALPFENADADAPAEILNHYSEIYQONIEKRISSETGNRTDMLRTHLHDFVTNPNMKNKDSN 931
Dy 376 IDNRQMFSLKDNQAMLNIACTIVQQAIEIKIGVANSTAWLKTATDIFRTQPOWNMSSE 435
Qy 932 VNFSGIFOGGFLKYENSOLTPYANSDYRLGLRMPIN-----IKDQYRGGEFLLAND 984
Dy 436 P--KNHDLQNGALTFVNSPLTPDNTNSFRLLNPTPTNQTGVPXYITDQSKGGFELLAND 493
Qy 985 IDNSNPVQAEQLNWLNYLYNFGTITANNDQANFSDVRVDAPNDIDADLMNIAQDYFNAA 1044
Dy 494 VDNSNPVQAEQLNWLNYLYNFGSITANDSAANFDGIRVDVNDVADLLOIAADYFKA 553
```



Db 849 FD-----NTEIYGRNPOVSGFLAVWVPGADSHQDARTLSDDTAHHDGKTFH 896  
QY 1401 TNAAFDSNVIEAFSNFYPTKESERANVRIAONADFEASLGFTSFEMAPQYNSSKDR 1460  
Db 897 SNAALDSQVIEGHSNFOAFATNTEDYNAVIAKNGOLFQDWGITSQLAPOFRSSTDS 956  
QY 1461 FLDSTINDGAFTRDYDLGMSPEKNGYGTDEDLNAIQALHAGLQVNMADWVPQIYNLPG 1520  
Db 957 FLDSTIQNGAFTRDYDLGMSPEKNGYGTDEDLNAIQALHAGLQVNMADWVPQIYNLPG 1016  
QY 1521 KEVATVTRDDRGVNMWDAIINNLLYVNTIGGGEYKQYGGAFGLDKLQKLYPEITKKQ 1580  
Db 1017 QELATVTRNSYGDKNDSIDQQLYVQSGGGKYQAQYGGAFGLDQIKQYFALFETQ 1076  
QY 1581 VSTGVAIDPSOKITEWSAKYFNGFNHILHSGVYVLKADG-QGYNL--GTTTKQFPIQL 1637  
Db 1077 ISTGLPMDPSOKITEWSKGFNGSNIQKQAGYVLKDSGTDOYKYVTSNNNRDELFPKQL 1136  
QY 1638 TGEKKGNGEGRVKGNDGNYFYDLAGNMVKNFTFEDSVGNWYFDDQDGKVENKHFVDVD 1697  
Db 1137 TDDLSETR--FVRDNIQWVY-YTUSGLARNTFIQDDNGNYFYFDSGTGHLVTG--FONIN 1191  
QY 1698 SYGKGEYFFLKNQVSPRGLVQ-TDNGTYFYFDNYGKVRNQTI-----NAGAM 1745  
Db 1192 NH----HYFFLPNGIELVQSFQONADGSTIYFDQKGRQVFNQYITDGTGYFYFQNDGT 1247  
QY 1746 IYT-----LD-----ENGLIKASYNDAE 1765  
Db 1248 VTSFTFIDGHKQFYKNGTQVRKQFVSDTD 1278  
RESULT 6  
Q54178 PRELIMINARY; PRT; 1577 AA.  
ID AC Q54178: Q54247;  
AD Q54178: Q54247;  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DI 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
GN GTFG.  
OS Streptococcus gordonii Challis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=29390;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=CHALLIS;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vickerman M.M.; Sulavik M.C.; Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
phase variants."; Dev. Biol. Stand. 85:309-314(1995).  
RL [2]  
RN SEQUENCE OF 1-96 FROM N.A.  
RP STRAIN=CHALLIS;  
RC MEDLINE=92276337; PubMed=1534326;  
RX Sulavik M.C.; Tardif G.; Clewell D.B.;  
RT "Identification of a gene, f99, which regulates expression of  
glucosyltransferase and influences the Spp phenotype of Streptococcus  
gordonii Challis."; J. Bacteriol. 174:3577-3586(1992).  
RL J. Bacteriol. 174:3577-3586(1992).  
DR EMBL; U12643; AAC43483.1; -;  
DR EMBL; M89776; AAA26969.1; -;  
DR InterPro; IPR002479; CW\_Binding.  
DR InterPro; IPR003318; GH\_70.  
DR Pfam; PF01473; CW\_Binding\_1; 18.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW transferase.  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 25.2%; Score 2375.5; DB 2; Length 1577;  
Best Local Similarity 47.4%; Pred. No. 9.4e-87;  
Matches 493; Conservative 159; Mismatches 328; Indels 59; Gaps 18;

QY 738 LQTINQOQYIIDPTTQCPKRNFLQSGNNWYIFDSGTGVTNALELOPAKGTVSSNEQYR 797  
Db 228 IKTIDGKKYIVQ--DGTGVKKNFVLENGKILYFDAETGALVDSYFOQOQTSLSNNEFT 286  
QY 798 NGNAAYSYDKSKSTENYNGYLTADTWYRPQKILKDGSTTWTDSKETDMRPILMYWWTNTLQ 857  
Db 287 OKNAFYGTDDKDIETVDGYLTADSNWYRPKFLKDGKTWTASTETDLRPLMLAWWDPDKRTQ 346  
QY 858 AYLNLNKKOHGNLLSALPFFNADADPAELNHYSEIVQONIEKRISEGTNTDLTLMHD 917  
Db 347 IYLNLMNGBE-NL---GIGAFESKTEQVLLTNVAQOVQKRIERISKEGDTKWLRTLSA 402  
QY 918 FVTNPMNR--KDSNVNPFSGTFOGGFKYENSGLTPYANSDYRLGLGMPN----- 968  
Db 403 EYKTPQNNWIKTESETTGKTKHLOGGALLYTNSDKTSHANSRYLLNRTPTTSQTGPKY 462  
QY 969 IKDQYRGGBEFLANDIDNSNPVQAEQNLWLYLLNFGTITANNDQANFDSVRVADPDN 1028  
Db 463 FDKSNGSYEFLLANDFDFNSPVAQAEQNLWLYLLNFGTITANNDQANFDSVRVADPDN 522  
QY 1029 IDADLMNIAQDYFNAAYGM--DSDAVSNKHINILEDWNHADPEYFNKIGNPOLTMDTITKN 1087  
Db 523 VNADLLQIASDYFKSPRYKVGSEEEALKHLSILEAWSDNDPDYNDKTKGAQLAIDNKRL 582  
QY 1088 SLNHG--LSDATNRWGLDAIVHOSLADRENNSTENVIPNYSPVRAHDNNSQDQIONAIRD 1146  
Db 583 SLLYSFMKRLSIRSGVEPTITNSLDRSTENKGBERTANIYFVRAHDSVQTVIADIIRE 642  
QY 1147 --VTGKDYHTFTFEDEQKIDAYIODNSTYKKNLYNIPASYAILLTNKDPITPVYIGD 1204  
Db 643 NINPNTDGLTFTMDLQKAFIYNEDMRKADKQYTFQNIPTAHALMLSNKDSITRYVYGD 702  
QY 1205 LYTDGGQYMEHOTRYDITLNLKSRVYAGGQSMQMSVG-----GNNTLTSVR 1256  
Db 703 LYTDGGQYMEKSPYHDAIDALLRARIKYVAGGQDMKYTMGVPYREADKWSYNGILTSVR 762  
QY 1257 YGKGAMTATDTGTCTGTGTGIGVYSNTPNLKLVNDKVLHMGAAHKNQYRAAVLTTT 1316  
Db 763 YGTGANEATDEGTAERTQGMVIAASNNPNLKLNEWDKLVNMGAAHKNQYRPVLLTTK 822  
QY 1317 DGVINWTSQGAAPVAM---TDENGDLYLSSHNLVYNGKEADTAVQGYANPDVSGYLAW 1373  
Db 823 DGISRYLTDEEYVPSQSLWKKTDANGILTDMND-----IAGYSRWQVSGYLAW 870  
QY 1374 VPVGASDNQDARTAPSTEKNSSNAYRTNAAFDSDNVIFEAFSNFYTTTKESERANVRIA 1433  
Db 871 VPVGAKENQDARTVASKKKNASQGYESSPALDSQLITTEGFSNFDQFATRDQYTNKVIA 930  
QY 1434 QNADFFASLGFTSFEMAPQYNSSKDRITFLDSTIDNGYAFTRDYDLGMSPEKNGYGTDEDLR 1493  
Db 931 KVNLFKEMGVTSFELPPQYVSSQDGTFLDSIIQNGYAFEDRYDMKSNKNKYGSLDOLL 990  
QY 1494 NAIQALHKGALQVMDWVPDQIYNLPKGEVATVTRVDDRGVNMWDAIINNLLYVNT-IG 1552  
Db 991 NALRALHVSNIQAIADWVPDQIYNLPKGEVATVTRVDDRGVNMWDAIINNLLYVNT-IG 1050  
QY 1553 GGEYKQYGGAFGLDKLQKLYPEITFKQVSTGVAIDPSOKITEWSAKYFNGTNIHRSYG 1612  
Db 1051 GTDYQKYGGAFLDELKAKYBEIPERVQISNGQKMTDEKTTKWSAKYFNGTNIHRSYG 1110  
QY 1613 YVLKADGGQYINLGTITKQFPIQLTGKKGKNGEGRVKGNDGNYFYFDLQAGNMYNFIE 1672  
Db 1111 YVLKDWGSKKEYLSNKNGETALPKQLV--NKEASTGFVKADTNG-FKFSYSTGSKAQDFFIQ 1167  
QY 1673 DSVGNWYFFDQDGKVENKHFVDVDSYGEKGYFFLKNQVSPRGLVQVTDNGT-YYPDNY 1731  
Db 1168 DENGWNYFVFNQGYLVGTGAREID---CKQ--LYFMKNGVQLRDALQDENGNQYYDKT 1221  
QY 1732 GKWVRNQINAGAMYITLD 1750  
Db 1222 GAKVLR-----YYTSD 1233



```
DR EMBL: L35495; AAC41412.1; -
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; GH_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E96A46 CRC64;

Query Match 25.2%; Score 2373.5; DB 2; Length 1449;
Best Local Similarity 41.5%; Pred. No. 1e-86;
Matches 560; Conservative 180; Mismatches 435; Indels 175; Gaps 42;

QY 491 ITMT-SNGHLAGWASDMSINETTPYAILLNGKEVTRKMSLT-ARPDVAAYVPSLYNS 549
DB 17 VTIATVGL-----SLGASVSLGTNDGVVQADEHTDAVAIPDTVDTGTVSND 66
QY 550 AVSGFD--TTIKLNDYQALNGLOVLLRFSSKAADNPSGDNVTVDQFSKNYAITGNN- 606
DB 67 TTAQODPTTAAATND---VATDQATPTATPDLTTDTT---NTVAANAADVTATVGTDR 119
QY 607 ---FDYKVGNGVVERSGWHATNOSNDKOSOWIILVNGKEVKROLVNDTKEGAAGFNR 662
DB 120 AATINDTATNDTAVD-----TTNNNTTDTI---TTVDRAATTERRATGARPGTGRR 171
QY 663 NDYKVNPAIENSMSFGQIITLPVTYKNEVQLVHRFSNDVKTEGNYVDFWSELMPV 722
DB 172 --ATPVNGTNTNANNNT-----VTVNNDLPA---TNNVVT----- 201
QY 723 KDSQKGNGLKQGLQTIQGYIIDPTTGPKNFLLQSGNNWIYFSDTGVGTNALE 782
DB 202 -----DGPSH---IKTIKGQYIYE--DDGIRKNYVLERIGGSQYFNAETGELSNOKE 250
QY 783 LOFAK--GTVSNEQYR-----NG--NAAYSYDDKSIENVNGYLTADTYWRPKOILKDG 833
DB 251 YRFDKNGGTGSADSTNTNTVNGDKNAFYGTDTDKIELVDGYFTANTWYRKPKEILKDGK 310
QY 834 TWDSKETDMPILVMWPNLTLOAYLYNMKOHNLPLSALPFENADA--DPAELNHS 891
DB 311 EWTASTENDKPLLVWMPSSKAIQASLYNMKEOGLTNQYTSFSOTOMDQAALE--- 367
QY 892 EIVQONIEKRISERTGNTDMLRTLMEDFVTNNPMNKNDSNVNFGIOFGGFLKYENS 951
DB 368 --VOKRIERIAREGNTDMLRTTIKFNKVTOPGNWSTSENLD--NNDHLOGGALLYNDSR 424
QY 952 TPYANSYRLLGRPIN-----IKDQYRGQEFLLANDIDNSNPVQAEOLNWLHY 1002
DB 425 TSHANSDFRLLNRPTSGTGHKHPKYTKDTSNGGFEFLANDIDNSNPVQAEOLNWLHY 484
QY 1003 LLNFGTITANNDQANFDSVRVDAPNDIDADLNIAQDYFNAAYGMD--SDAVSNKHINILE 1061
DB 485 IMNIGTITGGSEDEDFDGVVDVADVNDVNDLLOIASDYFRKAKYGADQSDQAIKHLILE 544
QY 1062 DNHADPFFYFNKIGNPQLTMDDTTKNSLNHG--LSDATNRGLDAIVHOSLADRENNTEN 1120
DB 545 AWSHNDAYNEDTKGAQLPMDPDPHMLALVYLLRPIGNRSGVEPLIENSLNDRSESKNS 604
QY 1121 VVIPNYSFVRADNNSDOICNAIRD-----VTGKDYHTFFEDQKIDAYIQDQNSTV 1175
DB 605 KEMANYAEVRAEDSEVSGIIOIKINEINPQSTG---NTFLDEMKKAFELYNKDMRSAN 661
QY 1176 KKNLYNIPASAYAILLNKNDTIPRYXYGDLYTDGGQYMEHQTRYDITITLLKSRVKRYA 1235
DB 662 KQYQYINPSAYALMLTHKDVPRVYGYDMYTDGQYMAOKSPYDAIETLLKGRIRYAA 721
QY 1236 GGQSMQTSVSG--GNNN-----ILTSVRYGKAMATATDGTDETRTQOGIGVYVSNTPNLK 1288
DB 722 GGQDMKVNVIQYGTNGWDAAGVLTISVRYGTGANSASDTGTAEFRNCGMAVIVSNQPALR 781
QY 1289 LGVNDKVVLLHMGAAHKNOQYRAAVALTTTDDGVINTSDQAGP---VAMTDENGDLVLSHNL 1346
DB 782 LTSN--LTINNGAHRNQAYRPLLTINDGVATVILNDSANGIVKYITDGNGLTFSANE- 838
```

```
QY 1347 VVNGKEEATAVOGYANPDVSGYLAWVPVVGASDNDQARTAPSTBKNSGNSAYRTNAAPD 1406
DB 839 -----IRGRNPQVDGYLAWVPVVGASENQDVRVAPSKKNSGLVYESNAALD 887
QY 1407 SNVIFEAFSNFVTPTKESERANVRIAQNADEFASLGFTSFEMAPQYNSKSKDRTFLDSTI 1466
DB 888 SQVIYSGFNFQDFVONPQYINKKIAENANLFKSWGITSFEFAQYYSDDGSLDSVI 947
QY 1467 DNGYATDRIYDGLGMSNPKNKYGTDEDLRNAIQALHRAGLQVMDVDPQIYNLPGREVATV 1526
DB 948 QNGYATDRIYDGLGMSKDNKYSLADLKAALKSHUAVIGISAIADVDPQIYNLPGDEVTA 1007
QY 1527 TRYDDRGNYWKDAIINNLIYVYNT--IGGGEYOKKYGGAFDLKLOLYPEIFTKKOVSTGV 1585
DB 1008 TRVNNYGETKDGALIDHSLYAAKTRTFGNDYOGKYGGAFDLKELKRLYPOIFDVOISTGK 1067
QY 1586 AIDPSOKITREWSAKYPNGNIIHRGSGYVLKADGGYYNLGTT--TKOFIPIOLTOKEKKG 1644
DB 1068 RMTTDEKITQWSAKYMNGTNILDGSEYVLKNGLNGY--GTNGGKVSLLP--KVVG--SNOS 1123
QY 1645 NEGFVKGNDG-----YFYDLAGNMVKNTFIEDSGVNWYFFDQDQKMWENKHFV 1694
DB 1124 TNGDNONGSGSGKFEKRLFSVRYRN--NQYAKNAFIKNDGNYFYFDNSGRMAYGEKTI 1182
QY 1695 DVDSYGEKTYTFLKNGVSFRGLVQTDNG--TYFYDNYGKM-----VR 1736
DB 1183 D-----GKQ--YFFLANGVOLRDGYRQNRGQVFFYDQGLVLANGKQDKPKDNNNASGR 1236
QY 1737 NOTINAGAMYT--LDENGLKLIKASYNDAE 1765
DB 1237 NQFVQIGNVWAYDNGRVRVTHQNINGQ 1266

RESULT 9
ID O69397 PRELIMINARY; PRT; 1455 AA.
AC O69397;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glucosyltransferase-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_Taxid=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4457;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hanada S.,
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA36120.1; -
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162914 MW; A1263427BF24E8E1 CRC64;

Query Match 24.9%; Score 2346.5; DB 2; Length 1455;
Best Local Similarity 41.0%; Pred. No. 1.2e-85;
Matches 536; Conservative 188; Mismatches 397; Indels 187; Gaps 37;

QY 549 SAVSGFTTIKLTNDQYQALNGOLOVLLRFSKA-----ADGNPSG---DN 590
DB 29 TSLGSLVKADSTDDRQQAFTESQASLVTTSEAKETLTATDTSTATSATQATVTDN 88
QY 591 TVTDQSKNATGGNFDVKNVNGVVERSGWHATNOSNDKOSOWIILVNGKEVKROLV 650
```



Db 89 VSTINQSTNTANTANFVDKPTTISE-----QAKTNSDK-----IIATSKAVNRLT- 135  
QY 651 NDRKEGAGNRNDVYKYNPAIENSMGSGQGIITLPIVIVKNEVQVLRHFSNDVKTEG 710  
Db 136 -----ATG-----KFPANNNTAH-----PKVTVD----- 155  
QY 711 NYVDFWSELMPVKDSFOKGGKPLKQ-----FGLQTINGQQOYIIDPTTGQP 755  
Db 156 -----KIVIPKPKI-----GKLKQPSLSQSDDIAALGNVKNIRKYNKYKYYY-KEDGTL 203  
QY 756 RKNFLQSGNNWYFDSOTGVGNALBQFAKTVSSNE---QYRNGNAAYSYDDKSIE 812  
Db 204 QKNYALNINKTEFFDETGTALSNNLTFSK--KGNITNNDNTNSFAQYQVYSTDVANFE 261  
QY 813 VNGYLTADTWYRKQILKDGTTWDSKETDMRPILMYWNPNTLTQAYLYNMYKQHGNLLP 872  
Db 262 VDHLYTAESWYRKYILKDKGKTWTQSTEKDFRPLMTWPDQETQOQYVYNNMAQLGIHQ 321  
QY 873 SALPFFNADAPAEHLNHYSEIVQONIEKRISETGNTDMLRTHDFVNNHMPNKNKSENV 932  
Db 322 T-----YNTATSPQLNLAQTQIKIEKITAENKTNWLRQTISAFVKTSANNSDEK 377  
QY 933 NFSGIQGGFLXYENSDLPYANSYRLLGRMPIN-----IKDQYRQGEFLAN 983  
Db 378 FDDHLQ--KGALLYSNNKLTSSQANSNRYILNRTPTNGTKKDPRYADRTIGGYEFLAN 436  
QY 984 DIDNSNPVQAEQLNWLIIYLLNFCTITANNNDQANFDSVRVDADPNIDADLMNIAQDYFNA 1043  
Db 437 DVDNSNPVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDADVNDVADLQLAGDYLKA 496  
QY 1044 AYGM-DSDAVSNKHINLEDNWHDADPEYFNKIGNPQLTMDTDTKNSLNHLSDATN-RWG 1101  
Db 497 AKGIHKNDKAANDHLSLEAWSYNDTPLYLHDDGDMNIMDNRLSLYSLAKPLNORSG 556  
QY 1102 LDALVHOSLADRENNSTENVIPNYSVVRADNNSQOIQNAIR-----DVTGKDYHFT 1156  
Db 557 MNPLITSLNVRTDNEATAVSPYSFIRAHDSVQDLIRLKAENPNVVG---YSET 613  
QY 1157 PEDSOKGIDAYIQONSTVKYNNLYNPASVAILLTKNDKIPRYVYGLDLYTDGGQYMEHQ 1216  
Db 614 MEEIKKAFEIYNKOLLATKRYTHYNALSYALLTNKSSVPRVYVYGMDFTDGQYMAHK 673  
QY 1217 TRYDYTLNLKSRVKVAGQSMQTSVGGNNILSVRYGKAMTADTGTDETRTQG 1276  
Db 674 TINYEAETLLKARIKIVSGQAMRNQV--GNSEIISVRYGKALKATDGTGTRTSG 732  
QY 1277 IGVVVSNTPNKLGYNKVKVILHMGAAHKNQOYRAVLITTTDGVNYSQDQAP---VAMTD 1334  
Db 733 VAVIEGNNPSRLKASDRVYVYMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTN 792  
QY 1335 ENGDIYLSHNLVYNGKEEADTAQVGVANPDVSGYLAWVPVGAASDNDQARTAPSTEKNS 1394  
Db 793 DRGELIFTA-----AD--IKGYANPOVSGYLGWVPVGAADQDVRAASTAPST 840  
QY 1395 GNSAYRTNAFDSNVIFEAFNSFYVPTPKESERANVRIAQNADFFASLGFTSFEMAPOYN 1454  
Db 841 DGKSVHQNALDSRVWFEGFSNFOAFATKKEEYTNVVIKAKNVDFAEWGVTFDEMAPOYV 900  
QY 1455 SSKDRTFLDSIDNGYATDSDYDGLGMSFPNKYGTDEDLRNAIQALHRAKLOVMADWVPDQ 1514  
Db 901 SSTDGSFSDVQNGYATDSDYDGLGMSFPNKYGTDEDLRNAIQALHRAKLOVMADWVPDQ 960  
QY 1515 IYNLPGKEVATVTRVDDRGVYWKDAIINNLYVYN--TTGGGEYQKYGGAFLDKLQKLYP 1573  
Db 961 MYALPEKEVWATRVYDKYGTTPVAGSQIKNTLYVVDGKSSGKDQAKYGGAFLEELQAKYP 1020  
QY 1574 EIFTKQVSTGVAIDPSOKITWESAKYFNGNINLHRSYGVYLVKADGGQYNNLGTHTTKQFL 1633  
Db 1021 ELFAKQISTGVPMDDPSYKIKQWSAKYFNGNINLGRGAGYLVKQCATNTYFSLYSDNTFL 1080  
QY 1634 PIQLTGERKQGNEGEVKG--NDG-NYFEYDLAGNMVKNFTFEDSVG--NWFFDQOGKAVE 1689

Db 1081 PKSLV-NPNHGTSSSVTGLVFDGKGYVYVYSTSGNOAKNAFI--SLGNWYFDNNGYMT 1137  
QY 1690 NKHFVDVDSYGEKGYFFFLKNGVSFRGVLQV-----DNGTY-----Y 1727  
Db 1138 GAQSN-----GANYFSLNSGILQIRNAILYDNGKNVLSYVYNGDGRYENGYLFGQWRY 1191  
QY 1728 FDNYGKMYRNQNTINAGAMITYLTDENGKLIKASYNDAEYPTSTVDGKM 1775  
Db 1192 FQN-GIMAVGLTRVHGAVQY-FDASG-----FQAKGQFITTAD-GKL 1230  
RESULT 10  
O69388 PRELIMINARY; PRT: 1455 AA.  
AC O69388;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Glucosyltransferase-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1309;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-WT4239;  
RX MEDLINE=98231643; PubMed=9370124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88638; BAA26110.1;  
DR InterPro; IPR002479; CW\_Binding.  
DR InterPro; IPR003318; GH\_70.  
DR Pfam; PF01473; CW\_Binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163046 MW; 6D90A4978D35D82 CRC64;

Query Match 24.9%; Score 2343.5; DB 2; Length 1455;  
Best Local Similarity 41.1%; Pred. No. 1.6e-85;  
Matches 534; Conservative 187; Mismatches 409; Indels 169; Gaps 35;  
QY 549 SAYSGVDTTIKLNDQYQALNGQLVLLRFSKAADGNPSGDNT-----VTDQFSKN 599  
Db 29 TSLSGSLVRADSTDDDRQAVTESQASLVTSEAKETLTATDTSTATSATSQLTATVDN 88  
QY 600 YATGGNFDYVYKNGNOVEFSGWHATNQSDKDSQMIIVLVNGKEVKRQLVNDTKEGAG 659  
Db 89 VSTTNQS---TNTANTANFVDKPTTSEOSKTDNSDKIATSKAYNRLT-----ATG 138  
QY 660 FNRNDYVKNPAIENSMGSGQGIITLPIVIVKNEVQVLRHFSNDVKTEGNYVDFWSEL 719  
Db 139 -----KFPANNNTAH-----PKVTVD-----KI 157  
QY 720 MPVKDSFQKGGKPLKQ-----FGLQTINGQQOYIIDPTTGQPRKNFLQSG 764  
Db 158 VPIKPKI---GKLKQPSLSQSDDIAALGNVKNIRKYNKYKYYY-KEDGTQKNYALNIN 212  
QY 765 NNWYFDSGTGVTGNALBQFAKTVSSNE---QYRNGNAAYSYDDKSIEVNGYLTADT 821  
Db 213 GKTFEDETGAISNNLTFSK--KGNITNNDNTNSFAQYQVYSTDAENFEHVHYLTAES 270  
QY 822 WYRPKQILKDGTTWDSKETDMRPILMYWNPNTLTQAYLYNMYKQHGNLLPSALPFFNAD 881  
Db 271 WYRPKYILKDGKTWTQSTEKDFRPLMTWPDQETQOQYVYNNMAQLGIHQT---YNTA 326  
QY 882 ADPAELNHYSEIVQONIEKRISETGNTDMLRTHDFVNNHMPNKNKDSNVNNGSGTQFC 941  
Db 327 TSPQLNLAQTQIKIEKITAENKTNWLRQTISAFVKTSANNSDEKPFDDHLQ-KG 385



```
QY 942 GFLKYENSLDTPYANSYRLILGRMPIN-----IKDQTYRGOELLANDIDNSPVY 992
Db 942 GFLKYENSLDTPYANSYRLILGRMPIN-----IKDQTYRGOELLANDIDNSPVY 992
QY 386 ALLYSNNKSLTQANSYRILNRTPTNQTKKDPRTADRTIGGYEFLANDVDNSPVY 445
Db 386 ALLYSNNKSLTQANSYRILNRTPTNQTKKDPRTADRTIGGYEFLANDVDNSPVY 445
QY 993 QAEQLNWLFLYLLNFGITITANDQANFDSVRDAPDNIDADLMNIAQDYFNAAYGM-DSDA 1051
Db 993 QAEQLNWLFLYLLNFGITITANDQANFDSVRDAPDNIDADLMNIAQDYFNAAYGM-DSDA 1051
QY 446 QAEQLNWLFLHFMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAAGKIHKNK 505
Db 446 QAEQLNWLFLHFMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAAGKIHKNK 505
QY 1052 VSNKHINLEDWHDPEYENKIGNPOLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSL 1110
Db 1052 VSNKHINLEDWHDPEYENKIGNPOLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSL 1110
QY 506 AANDHLSILEAWSYNDTPYJHDDGDNMINMDNRLRSLYSLAKPLNQRSGMPLITNSL 565
Db 506 AANDHLSILEAWSYNDTPYJHDDGDNMINMDNRLRSLYSLAKPLNQRSGMPLITNSL 565
QY 1111 ADRENNTENWIPNYSFVRAHDNNSQDQIONAIR-----DVTGKDYHTFTFEDEQKID 1165
Db 1111 ADRENNTENWIPNYSFVRAHDNNSQDQIONAIR-----DVTGKDYHTFTFEDEQKID 1165
QY 566 VNRDDNAETAAPSYSFIRAHSEVQDLIRNIARINENNVVG---YSTMEIEKKAFE 622
Db 566 VNRDDNAETAAPSYSFIRAHSEVQDLIRNIARINENNVVG---YSTMEIEKKAFE 622
QY 1166 AYIDQNSYVKKYNLYNIPASYAILLTKOTIPRVYGDLYTDGGYMEHQTRYDYTLN 1225
Db 1166 AYIDQNSYVKKYNLYNIPASYAILLTKOTIPRVYGDLYTDGGYMEHQTRYDYTLN 1225
QY 623 IYNKDLATEKKYTHYNTALSYALLTNKSVPRVYGMFTDGGQYMAHKTINYEAET 682
Db 623 IYNKDLATEKKYTHYNTALSYALLTNKSVPRVYGMFTDGGQYMAHKTINYEAET 682
QY 1226 LKSRVYVAGGQSMQMSVGGNNILTSVRYGKGAMTATDGTDETRTQIGVYVSNTP 1285
Db 1226 LKSRVYVAGGQSMQMSVGGNNILTSVRYGKGAMTATDGTDETRTQIGVYVSNTP 1285
QY 683 LKARIKYVSGQAMRQV--GNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNP 741
Db 683 LKARIKYVSGQAMRQV--GNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNP 741
QY 1286 NLKLGVDKVVYHMGAAHKNQOYRAAVLTITDGVINTSQAP--VAMTDENGDLVSS 1343
Db 1286 NLKLGVDKVVYHMGAAHKNQOYRAAVLTITDGVINTSQAP--VAMTDENGDLVSS 1343
QY 742 SLRLKASDRVYVYNGAAHKNQAYRPLLTDDNGIKAYHSDOEAAGLVRYINDRGELIFTA 801
Db 742 SLRLKASDRVYVYNGAAHKNQAYRPLLTDDNGIKAYHSDOEAAGLVRYINDRGELIFTA 801
QY 1344 HNLVYNGKEEADTAQVYANPDVSGYLAVVYVPGASQNDQARTAPSEKNSGSAVETNA 1403
Db 1344 HNLVYNGKEEADTAQVYANPDVSGYLAVVYVPGASQNDQARTAPSEKNSGSAVETNA 1403
QY 802 -----AD--IKGYANFQVSGYLGWVVPVGAARDQVRYAASAPSTDGKSVHQA 849
Db 802 -----AD--IKGYANFQVSGYLGWVVPVGAARDQVRYAASAPSTDGKSVHQA 849
QY 1404 AFDNSNVIPEAPSNVYPTKESERANVRIAQNADFFASLGFTSFEMAPQNSKDRFTFLD 1463
Db 1404 AFDNSNVIPEAPSNVYPTKESERANVRIAQNADFFASLGFTSFEMAPQNSKDRFTFLD 1463
QY 850 ALDSRVMEFEGSNFQAFATKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSSTDGSLD 909
Db 850 ALDSRVMEFEGSNFQAFATKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSSTDGSLD 909
QY 1464 STINGVAFTRDYDLGSEPNKYGTDDELNAIQALHAGLVQVWADWVPQIYNLPQKEV 1523
Db 1464 STINGVAFTRDYDLGSEPNKYGTDDELNAIQALHAGLVQVWADWVPQIYNLPQKEV 1523
QY 910 SVIQNGVAFTRDYDLGSEPNKYGTDDELNAIQALHAGLVQVWADWVPQIYNLPQKEV 969
Db 910 SVIQNGVAFTRDYDLGSEPNKYGTDDELNAIQALHAGLVQVWADWVPQIYNLPQKEV 969
QY 1524 ATVTRVDRGNVWDAIINNLYVN--TIGGEYOKYKGAFDLKLOKLYPEIFTKQVS 1582
Db 1524 ATVTRVDRGNVWDAIINNLYVN--TIGGEYOKYKGAFDLKLOKLYPEIFTKQVS 1582
QY 970 VTATRVKYGTVPVAGSQIKNTLYVVDGKSGKQDQAKYGGAFLELOAKYPELFARKQIS 1029
Db 970 VTATRVKYGTVPVAGSQIKNTLYVVDGKSGKQDQAKYGGAFLELOAKYPELFARKQIS 1029
QY 1593 TGVAIDPSQKITESSAKYFNTNLRHSGVYVLKADGGQYVNLGTTTKQLPIQLTGEKK 1642
Db 1593 TGVAIDPSQKITESSAKYFNTNLRHSGVYVLKADGGQYVNLGTTTKQLPIQLTGEKK 1642
QY 1030 TGVMPDPSVKIKQWSAKYFNTNLRHSGVYVLKADGGQYVNLGTTTKQLPIQLTGEKK 1088
Db 1030 TGVMPDPSVKIKQWSAKYFNTNLRHSGVYVLKADGGQYVNLGTTTKQLPIQLTGEKK 1088
QY 1643 QGNEGFYKVG--NDG--NYFYDLAGNVKNTFIEDSVG--NMVFFPDQDGKMYENKHFVDVDS 1698
Db 1643 QGNEGFYKVG--NDG--NYFYDLAGNVKNTFIEDSVG--NMVFFPDQDGKMYENKHFVDVDS 1698
QY 1089 HGTSSSVTGLVFDGKGVYVYSTSGVQAKNTFI--SLGNMWTYFDNNGYVMTGAQSIN-- 1143
Db 1089 HGTSSSVTGLVFDGKGVYVYSTSGVQAKNTFI--SLGNMWTYFDNNGYVMTGAQSIN-- 1143
QY 1699 YGKGTYPFLKNGVSFRGLVQT-----DNGTY-----YFDNYGKMVR 1736
Db 1699 YGKGTYPFLKNGVSFRGLVQT-----DNGTY-----YFDNYGKMVR 1736
QY 1144 ---GANYVFLNGIQLRNAIYDNGNKVLSYVNGDGRYVNGYVYLFQGOVRYFQN-GIMAV 1199
Db 1144 ---GANYVFLNGIQLRNAIYDNGNKVLSYVNGDGRYVNGYVYLFQGOVRYFQN-GIMAV 1199
QY 1737 NOTINAGAMITLDRNGKLIRKASVNSDAEYPTSTDVGRM 1775
Db 1737 NOTINAGAMITLDRNGKLIRKASVNSDAEYPTSTDVGRM 1775
QY 1200 GLTRYHGVAVQY-FDASG-----FQAKGFITTAD-GLK 1230
Db 1200 GLTRYHGVAVQY-FDASG-----FQAKGFITTAD-GLK 1230
RESULT 11
O69382 ID O69382 PRELIMINARY; PRT; 1455 AA.
AC O69382;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Glucosyltransferase-SI.
GN GTF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
```

```
OX NCBL_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RA "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL PEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26102.1;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162970 MW; 27D4D3A1EECA2939 CRC64;
Query Match 24.9%; Score 2342.5; DB 2; Length 1455;
Best Local Similarity 41.4%; Pred. No. 1.7e-85;
Matches 535; Conservative 189; Mismatches 412; Indels 157; Gaps 35;
QY 549 SAVSGPDTTIKLTNQYQALNGQLQVLLRFSKA-----ADGNPSG---DN 590
Db 29 TSLSGSLVKADSTDDRQQAQVTSQASLVTTSEAAKETLTATDTSTATSQPTAIVTDN 88
QY 591 TVTDQFSKNVATGGNFDYKVGNGQVSESGHATNQNDKDSQWIIYLVNGKEVKRQLV 650
Db 89 VSTINQSTNTTANTANFDVKTPTTSE-----QSKTDSK-----IATSKAVNLT- 135
QY 651 NDTKEGAAQFNNDVYKVNPAJENSSMSGFQGIITLPVTVKNNVQLVHRFSNDYKTGEG 710
Db 136 -----ATG-----KFPANNNTAHSRTVDKVIKPKIGKLPSSLDQDDIAALG 182
QY 711 NYVDFWSELPVKDSQKNGPLKQGLQTINGQQYYIDPTTQGPKNFLLSGNNWYIF 770
Db 183 N-----VKN-----IRKNGKYYY-KEDGTLOKYNALINAKTFF 218
QY 771 DSDTGVGTNALEQPAKGTVSSNE---QVRNGNAAYSYDKSINENYGLVLTADTYWRPKQ 827
Db 219 DETGALSNLTLPK--KGNITNDNTNSAQYNOVYSTDAANFEHVDHVLTAESWRPKY 276
QY 828 ILKDGTTWTDKSETDMRPILMVWNPNTLQAYLYNTMKOHGNLLPSALFPFNADAPAL 887
Db 277 ILKDGKTWTQSTKEKFRLLMTWPDQETQRQVYVNMNAQLGIHOT---YNTATSPLO 332
QY 888 NHYSEIVQONIEKRISFTGTNTDMLRLMHDFVTNNPMNKNDSNVNFSGIFQGGFLKYE 947
Db 333 NLAQAQTIQTKIEKITAENNTNMLROTISAFVKTQSAWNSDEKPPDDHLQ-KGALLYSN 391
QY 948 NSDLTPYANSYRLILGRMPIN-----IKDQTYRGOELLANDIDNSPVYQAEQLN 998
Db 392 NSKLTQANSYRILNRTPTNQTKKDPRTADRTIGGYEFLANDVDNSPNVYQAEQLN 451
QY 999 WLYYLLNFGITITANDQANFDSVRDAPDNIDADLMNIAQDYFNAAYGM-DSDAVSNKH 1057
Db 452 WLHFLMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAAGKIHKNKKAANDHL 511
QY 1058 NILEDWHDPEYENKIGNPOLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSLADRENN 1116
Db 512 SILEANSYNDTPYLEDHDDGDNWKNKLNRLSLYSLAKPLNQRSGMPLITNSLVNRIDD 571
QY 1117 STENVVIPNYSFVRAHDNNSQDQIONAIR-----DVTGKDYHTFTFEDEQKIDAYIQDQ 1171
Db 572 NAETAAPSYSFIRAHSEVQDLIRNIARINENNVVG---YSTMEIEKKAFEIYNKDL 628
QY 1172 NSTVKKYNLYNIPASYAILLTKOTIPRVYGDLYTDGGYMEHQTRYDYTLNLKSRV 1231
Db 629 LATEKKYTHYNTALSYALLTNKSVPRVYVYGMFTDGGQYMAHKTINYEAETLLKARI 688
QY 1232 KYVAGGSMQMSVGGNNILTSVRYGKGAMTATDGTDETRTQIGVYVSNTPNKLGV 1291
Db 689 KYVSGGQAMRQV--GNSEIITSVRYGKGALKATDGTDRITRTSGVVYIEGNNPSRLKA 747
```





GN GTFR.  
 OS Streptococcus oralis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 CC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC10557;  
 RX MEDLINE=20231779; PubMed=10768934;  
 RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 encoding glucosyltransferase from Streptococcus oralis.";  
 RL Infect. Immun. 68:2475-2483(2000).  
 DR ENBL: AB025228; BAA95201.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; GH\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 17.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;  
 Query Match 24.7%; Score 2330; DB 2; Length 1575;  
 Best Local Similarity 42.6%; Pred. No. 6.1e-85;  
 Matches 506; Conservative 175; Mismatches 372; Indels 134; Gaps 23;  
 QY 624 ATGQNDKDSQWII--VLVNGKEVRQLVND--TREGAAGFRNRNDVYKYNPAIENSSMS 678  
 DB 161 AQDEGDKREKTAVEDKIVANPKVAKDRLPEPGSKOGAIA-ERMVADQAQPA----- 212  
 QY 679 GFGIITPLVTYKENVOLVHRSNDVKTGEGVNVDFWSELMVPKDSFQKGNGLPKQFGL 738  
 DB 213 -----PVNADHDDVLSH-----I 226  
 QY 739 QTINGQOYIDPTGQPRKFLQLSGNNWIFDSDTGVGTNALELOFAKTVSSNQYRN 798  
 DB 227 KTDGKNYVQ-DGTVKKNFAVELNGRIIFYDAETGALVDSNEYQFQOQTSILNFEFSQ 285  
 QY 799 GNAAYSDDKSIENVNGYLADTWYRPQILKDGTTWDSKETDMRPILMWMPNLTQA 858  
 DB 286 KNAEYGTGDKDIEYVDGILTADSWYRPKFLKDGKTWASTETDLRPLLMAMPDKRTQI 345  
 QY 859 YLYNMQHGNLPSALPFFNADADPRLNHYSEIVQONIEKRISEPTGNTDWRLLMHDF 918  
 DB 346 NYLYNMNQOQ---LGAGAFENKVEQALLTGASQOVORKIEEKIGKGDIKWLTLMGAF 401  
 QY 919 VTNNPMWN--KDSNVNFSQIFQGGFLKYENSDLTPTAYNSDYRLGLRMPIN-----I 969  
 DB 402 VKTOPNWNKIESETTGTCKDHOGGALLYTNNEKSPADSKFLLNRTPTSTGTGPKYF 461  
 QY 970 KDQYRGOEFLLANDIDNSNPVQAEOLNMLYLLNFGTITANNDOANFDSVRVADPDI 1029  
 DB 462 IDKNGGVEFLLANDFNSNPVQAEOLNMLYMNFGSIVANDPTANFGVRVDAVDNV 521  
 QY 1030 DADLMTAQQDYFNAAYGM-DSDAVSNKHINILEDWNHADPEYFNKIGNPOLTMDDTIKNS 1088  
 DB 522 NADLLQIASDYFKRYKVGSEEEAIKHLSTLEAWSNDPDYKNTDRTGAQLAIDNKLRLS 581  
 QY 1089 LNHG-LSDATNRGLDALVHOSLADRENNSTENVVINYFVRAHDNNSODIQNATR- 1146  
 DB 582 LLTFSMRLSRSGVEPTITNSLDRSEKKNGERMANYIFVRAHDSEVQTVIADI IREN 641  
 QY 1147 -VTGKDYHTTFEBOKGIDAYIQDQSTVKKYNLYNIPASYAILLNKDTIPRVYGD 1205  
 DB 642 INPNTDGLTFTMDLQKAFIYNEDMRKADKKYTFQFNPTAHALMLNSKDSITRVYGD 701  
 QY 1206 YTDGQYMEHETRYDILTLLKSRVYVAGQSQMOTMSVG-----GNNNLTISVRY 1257  
 DB 702 YTDGQYMEKSPYHDAIDALLRARIKYVAGQDMKYTMGVPREADKWSYNGILTISVRY 761  
 QY 1258 GKGAMTADTGTDTETRGIGVYVSNTPNLKGLVNDKVLHMGAAHKNQOYRAAVLTTD 1317  
 DB 762 GTGANEATDGTETRTGOMAVIASNPNLKLNEWDKLOVNWGAHKNQYRVPVLLTTKD 821

QY 1318 GVINYTSDGAPVAM---TDENGDLYLSSHNLVNGKEEADTAVOGYANPDVSGYLAWV 1374  
 DB 822 GYSLRILTDEEVPQSLWKKTDANGILTFDMND-----IAGYSNVQVSGYLAWV 869  
 QY 1375 PYGASDNQDARTAPSTEKNSGNSAYRTNAAFDSNVIFPAFSNVFYPTPKESERANVIAQ 1434  
 DB 870 PVGAKADQDARTTASKKKNASQGVYESSAALDSQLIYEGFSNFQDFATRDDQYTNKVIK 929  
 QY 1435 NADFFASLGFTSFEMAPQYNSSKORTFLDSTIDNGYATDRYDLGMSBPNKYGTDEDLRN 1494  
 DB 930 NVNLFKENGVTSEFLPPQVSSQDGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLKDLN 989  
 QY 1495 AIQALHFKAGLOVMADVPDQIYNLPKGEVATVTRVDDRGNVWKDAIINNLLYVNT-IGG 1553  
 DB 990 ALRALHNVNIQAIADWPDIYNLPKGEVATVTRVNNGTVREGAEIKEKLYVANSKTNE 1049  
 QY 1554 GEYQKRYGGAFLDKLOKLYPEFTKKQVSTGVAIDPSKITENSAKYENGNTILHRGSGY 1613  
 DB 1050 TDFQGYGGAFLDELKAKYPEIFERQVQISNGOKMTDEKIKWSAKYFNGTNILGRGAY 1109  
 QY 1614 VLKADGGQYVNLGTTTKQFLPTQLTGERKKQGNFVKNGDNGYFYFDLAGNMVKNFTFED 1673  
 DB 1110 VLKDMASNDYLTNRNGEIVLPKQLV--NKNSYTGFEVSDANGT-KFYSTSGYQAKNSFIOD 1166  
 QY 1674 SYGNMYFFDQDGKVMENKHFVDVDSYGEKGYT-FFLKNGVSFRGLVQTDNGT-YFFDNY 1731  
 DB 1167 ENGNMYFEDKRGYLYTGAHEID-----GKHVYFLKNGIQLRDSIREDENGNQYYDOT 1219  
 QY 1732 GKMYRNOTINA-----GAMIYTL-----DENGKLIK 1757  
 DB 1220 GAQVNLRYTTDQGNWRVFDAGKVMARGLVKIGDGGQFFDENGYQVK 1266  
 RESULT 15  
 Q55263 PRELIMINARY; PRT: 1590 AA.  
 ID Q55263  
 AC Q55263  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE GTF-1.  
 GN GLUCOSYLTRANSFERASE.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 CC Streptococcaceae; Streptococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33478;  
 RA Sato S.;  
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase  
 produced from Streptococcus sobrinus ATCC 33478.";  
 RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
 DR EMBL: D63570; BAA09792.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; GH\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 15.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3E2C6B4FD43 CRC64;  
 Query Match 24.3%; Score 2289; DB 2; Length 1590;  
 Best Local Similarity 40.8%; Pred. No. 2.6e-83;  
 Matches 532; Conservative 180; Mismatches 399; Indels 194; Gaps 37;  
 QY 540 AAYPSLYNSAVSGEFTTITKLTNDQYALNGQLVLLRFESKAADGNPSGNTVTDQ---- 595  
 DB 24 ATMLASALGASVASATD-TASDDSNQAV-----VTGDTTNQATDQTSIA- 69  
 QY 596 ---FSKNYATITGQNFYKVNQVFGSGHATNQSNKDSQMIIVLVNGKEVKROLVND- 652  
 DB 70 ATATSEQASASTDAATQASA-ABQTQ-----GTTASTDTAAQ---TTTANAEAKW----- 115  
 QY 653 TKEGAAGFRNDVYKYNPAIENSSMSFGQIITPLVTYKENVOLVHRSF---SNDVKTGB 709

Db 116 -----VPTENENQGFTEMLAEAKNVATAE 140  
Qy 710 GNYVDWSELMPVKDFQKNGPLQFGLQTINGOQYIIDPTTGQPRKNEFLLOSNNWLY 769  
Db 141 SD--SPPSDL-----AKSNKQVQDGKYYID-QDGNVKKNFPAVSGDKIYY 184  
Qy 770 FDSDTGV--GTNALELOFAGKTGVSSNEQ-YRNGNAAYSDDKSIENKYNGLTADTWYRK 826  
Db 185 FD-ETGAYKDTSKVDADKSSAVSONATIFAANNRAYSTAENFEAVDNYLTADSWYRK 243  
Qy 827 QILKDGTTWTDKETDMPILMWNPNTLOAYVLYNMYKHGNNLLPSALFFENADADPAE 886  
Db 244 SILKDGKTWTESGKDDPRLLMAWNPDETKRYVNF-----NLVYGIDKTYTAETSQAD 299  
Qy 887 LNHYSEIVQNIIEKRISETGNTDLRLTLMHDFVNNPMNKNDSNFSGIQFQGGFLKY 946  
Db 300 LTAABELVQARIEQKITTEQNTKWLREAISAFVKTQPNNGESEKPYDD--HLQNGALKF 357  
Qy 947 EN-SDLTPYANSYRLGRMPINIKQO-----TYR-----GOEFLANDIDNSNPVQA 994  
Db 358 DQSDLTPDQCSNRLNRLNRTPTNQTGSLDSRFTYNANDPLGGYEFFLLANDVDNSNPVQA 417  
Qy 995 EQLNWLHYLLNFGTITANNQANFSDYRVDAPDNIDADLMNIAQDYFNAAYGMD-SDAYS 1053  
Db 418 EQLNWLHYLLNFGSIYAKDADANFSDIRVDAVDNVDADLLQISSDYLKAAVGDKNKNA 477  
Qy 1054 NKHINILEDNHNADPEYFNKIGNPOLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSLAD 1112  
Db 478 NNHVSIVEASDNDPTVYLHDDGDNLMNDKFRLSMLWSLAKPLDKRSGNLPLIHNSLYD 537  
Qy 1113 RENNSTENVVIPNYSFVRAHDNNSQIQONAIR-DVTGKY-HTFTFEDBQKGIDAYIQD 1170  
Db 538 REVDDREVEIVPSYSPARAHSEVQDIIRDIKAEINPSFGYSFTQEEIDQAFKIYNED 597  
Qy 1171 QNSTVKYKYNLYNPASPAILLTKWDTIPRYVYGDLYTDGQYMEHOTRYIDTTLNLLKSR 1230  
Db 598 LKTKDKKYTHYNPLSYTLTLTKNGSIPRYVYGDYMTDQGYMANKTVNYDAIESLLKAR 657  
Qy 1231 VKYVAGQSQMTSVGGNNHILTSVRYGKGAMTATDGTGTGTGIGVYVVSNTPNLKLK 1290  
Db 658 MKYVSGGQAMQNYQI-NGEILTSVRYGKCALQSKDGADATRTSGVGVYMGQPNFSL- 715  
Qy 1291 VNDXVW-LHGAANKNOYRAAULTTTDGVNINTSDGAP-----VAMTDENGDLYLSSH 1345  
Db 716 -DGKVALNMGAAHNGEYRALMVSTKDGATYATDADASKAGLVKRTDENGILYFLNDD 774  
Qy 1346 LVNNGKEADTAVOGYANPDVSGYLAVWPVGASDNDODARTAPSTEKNKNSAVRTNAAF 1405  
Db 775 L-----KGVANPOVSGFLOVWPVGAADDQDIRVAASDRASTDGKSLHODAAM 822  
Qy 1406 DSNVIFAFSNFVYPTKESERANVRIAQNADEFASLGFTSFEMAPQYNSKORTFLDST 1465  
Db 823 DSRVMEGFSNFQSFATKEEYTNVIANNVKVFVSGITDFEMAPQYVSTDGQFLDSV 882  
Qy 1466 IDNGYATDTRYDLGMSBPNKYGTDEDLRNAIOALHAGLOVMAWDVDPDIYNLPKGKAVAT 1525  
Db 883 IQNGYATDTRYDLGMSKANKYGTADQLVKAHAGLKVMAWDVDPDQMYTPKQEVVT 942  
Qy 1526 VTRVDDRGNVWKDAIINNLYVNT-IGGGEYKQYGGAFLDKLQKLYPEIFTKQVSTG 1584  
Db 943 VTRTDKFGKPIAGSQINHSLYVDTKSGDDYQAKYGGAFDELKEKYPELFTKKQISTG 1002  
Qy 1585 VALDPSOKITWSAKYNGINILHRGSGYVLKAD-GQOYVNLGTTTKQFLPIQLTGEKKQ 1643  
Db 1003 QAIIDPSVKIKOWSAKYNGSILRGADYVLSQVSNKYFNWASDT-LFLPSSLGKQVVE 1061  
Qy 1644 GNEGFEVKGNDGNYFYD--LAGNNVKNTFIEDSVGNWFFDQDGKVENKHFVDVDSYGE 1701  
Db 1062 SGIRY---DGKGYIYNSATGDQVKASFITEA-GNLYYFGKDGVMYTGATIN----- 1110  
Qy 1702 KGTFFFLKNGVSPRG-----GLVQ 1720  
||||| : |

Db 1111 GANYFFLENGTALRNTIYTDAGNSHYANDGKRYENGYOQFGNDWRYFKDGNMAVGLTT 1170  
Qy 1721 TDNGTYVYFDNYGKMVRNOTI--NAGAMIYTLDENGKLIKASYNSD 1763  
Db 1171 VDGNYQYFDKDGVOAKDKLIIVTRDGKRVYFDQHNNGNAVNTNFIAD 1215  
Search completed: April 23, 2003, 15:09:40  
Job time : 101 secs